Augmenting State Capacity for Child Development: Experimental Evidence from India*

Alejandro J. Ganimian† New York University

Karthik Muralidharan‡ UC San Diego

Christopher R. Walters§ UC Berkeley

January 13, 2023

Abstract

We use a large-scale randomized experiment to study the impact of augmenting staffing in the world's largest public early childhood program: India's Integrated Child Development Services (ICDS). Adding a half-time worker doubled net preschool instructional time and led to 0.29σ and 0.46σ increases in math and language test scores after 18 months for children who remained enrolled in the program, and corresponding increases of 0.13σ and 0.10σ for the larger population of all children enrolled at baseline. Rates of child stunting and severe malnutrition were also lower in the treatment group for children who remained enrolled. A cost-benefit analysis suggests that the benefits of augmenting ICDS staffing are likely to significantly exceed its costs. Several features of our study setting and design suggest that these effects are likely to replicate even at larger scales of program implementation.

JEL codes: C93, I21, I22, I25

Keywords: preschool education, early childhood development, nutrition, malnutrition, stunting, ICDS, state capacity, staffing, public service delivery, India

*We thank the Government of Tamil Nadu especially S. Krishnan, L. Manivasan, and R. Kannan for their support for this project. We also thank Aparna Krishnan, Gautam Patel, Megha Pradhan, and Prakarsh Singh for their support at various stages of this project. We thank Larry Aber, Prashant Bharadwaj, Gordon Dahl, Jishnu Das, Purnima Menon, Jayashree Raghunandan, Mauricio Romero, Abhijeet Singh, Sharon Wolf, Hirokazu Yoshikawa, and several seminar participants for comments. Shruti Agarwal, Tanay Balantrapu, Aditi Bhowmick, Sharnic Djaker, Smit Gade, Nandini Gupta, Putul Gupta, Murali Mallikarjunan, Nihal Ranjit, Dhananjay Singh, Sivaranjani Sivamohan, and Anuja Venkatachalam provided excellent research assistance and field support. This study was registered with the AEA Trial Registry (RCT ID: AEARCTR-0001772). It was approved by the Institutional Review Boards of the University of California, Berkeley and San Diego and the Institute for Financial Management and Research. We thank the Government of Tamil Nadu and the Abdul Latif Jameel Poverty Action Lab (J-PAL) for funding this study. All views expressed are those of the authors and not of any of the institutions with which they are affiliated.

†NYU. alejandro.ganimian@nyu.edu.

‡UC San Diego, NBER, J-PAL, and BREAD. kamurali@ucsd.edu.

§UC Berkeley, NBER, J-PAL, and MIT-SEII. crwalters@econ.berkeley.edu.

1 Introduction

There is a broad consensus in the child-development literature that children's early-life interactions with parents and teachers have important consequences for cognitive and socio-emotional development (Shonkoff, Phillips, et al., 2000; Engle et al., 2007; Heckman and Mosso, 2014). However, much of the evidence on these benefits comes from small-scale experiments in unrepresentative samples and situations, and it has proven challenging to realize the gains from small-scale studies at larger scales of implementation (List, Suskind, and Supplee, 2021; List, 2021). Thus, a key input into policy decisions on whether to expand funding for early-childhood development programs is whether there is evidence of positive impact at scale. However, there is very little experimental evidence on easily implementable, and cost-effective ways of improving early childhood human development outcomes at scale.

This paper contributes to filling this evidence gap in the context of the largest early childhood development program in the world: the Government of India's Integrated Childhood Development Services (ICDS). ICDS caters to over 36 million 3-to-6-year-olds, and it provides a range of early childhood health and nutrition services as well as preschool education.[1] ICDS also serves another 46 million children in the 0-3 age range with supplemental nutrition and health services provided through home visitation programs. ICDS services are provided through 1.35 million *anganwadi* centers (AWCs) across India by *anganwadi* workers (AWWs). These services are provided free of charge and disproportionately cater to poor and vulnerable populations. Thus, the ICDS is the Government of India's primary vehicle for reaching tens of millions of socio-economically disadvantaged children, who are most likely to be in need of high-quality early childhood education and nutrition programs.

Despite its importance, ICDS has limited staffing and funding. In particular, a *single* worker (assisted by a helper for cooking and cleaning) is responsible for health, nutrition, home visitation, and preschool education, in addition to several administrative tasks. Further, AWWs are recruited locally and are paid roughly one-fourth the salary of civil-service primary-school teachers. The combination of low marginal cost and potential high returns suggest that adding a staff member to focus on preschool education may be a promising policy option for strengthening the ICDS. Doing so could increase instructional time and also free up time of the primary worker to focus on health and nutrition activities.

We study the impact of such an approach by conducting a large-scale randomized experiment across a sample of 320 AWCs in four districts representative of a population of 60 million people in the state of Tamil Nadu. Half of these centers were randomly selected to receive an extra facilitator focused on early childhood education. The facilitator was scheduled to work for half a day and was paid half the salary of a regular worker on a full-time shift.

[1] For comparison, the entire Head Start program in the US had 652,422 funded slots as of 2019, which is less than 2% of the coverage of the ICDS (NHSA, 2020).

The intervention was implemented by the Government of Tamil Nadu, including hiring and training of the facilitators. By studying a representative population and administering the program with the same protocols that would be used if it were to be scaled up, our study generates policy-relevant evidence on the likely impacts of implementing the intervention at scale (Muralidharan and Niehaus, 2017; List, Suskind, and Supplee, 2021; List, 2021).

Our primary outcomes of interest are test scores in math, language, and executive function. We measured these through independent tests conducted in AWCs, as well as through independent tests administered in a representative sample of households. We also collected child anthropometric data as a secondary outcome. Finally, we measured worker attendance, timeliness, and time use through unannounced visits to the centers. We present four sets of results on time use, education, nutrition, and cost-effectiveness respectively.

First, adding the facilitator decreased the center closure rate from 12.5 percent in the control group to 3.8 percent in the treatment group. On average, facilitators provided nearly an hour of daily preschool instruction in treated centers during the two-hour observation window. Workers in treated centers reduced time spent on preschool education, but they spent more time on health and nutrition tasks, and on completing administrative work (while being in class and supervising the instructional work of the facilitator). Adding the time spent across the worker and the facilitator, we find that the total time spent on preschool education *doubled* in treated centers (from 38 to 76 minutes per day). Treated centers also saw over 150% increases in total staff time spent on health and nutrition activities (from 6 to 16 minutes) and on administrative work (from 20 minutes to 55 minutes) during the observation window.

Second, in line with this increase in total instructional time, we find significant increases in children's learning levels. Eighteen months after baseline, children in treated centers scored 0.29, 0.46, and 0.18 standard deviations (σ) higher on independent tests in math, language, and executive function conducted in the AWCs ($p < 0.01$ for all three subjects). Overall, the intervention boosted a composite measure of learning across all three domains by 0.29σ ($p < 0.01$). The gains were broad-based and the treatment distribution first-order stochastically dominates the control distribution.

While there was no differential attrition in the center-based tests across treatment and control groups, the follow-up rate from baseline to endline was only 33%. This reflects a combination of children graduating out of preschool, moving to private preschools, families migrating, and irregular attendance. We therefore supplement the AWC-based outcomes with household-based measurements for a representative sample of the baseline cohort, achieving an 89% follow-up rate with no differential attrition. Treatment effects on the household tests are smaller but still significant. Children in treated centers scored 0.13σ higher in math ($p < 0.01$), 0.10σ higher in language ($p < 0.05$), 0.05σ higher in executive function (not significant), and 0.11σ higher on the composite learning measure ($p < 0.05$).

The smaller effects on household tests most likely reflect the fact that the children who were not in the AWC endline testing sample (but who were included in the household sample) had either moved out of AWCs or attended them infrequently. This is corroborated by our finding of similar treatment effects across household and center-based estimates for the common sample of children who took both tests. Moreover, treatment-on-the-treated effects obtained by scaling the household sample estimates by the share of children observed at the center at the endline are close to the AWC and common sample estimates. We therefore interpret the AWC estimates as reflecting treatment effects on children who actively attended the centers, while the household estimates capture intent-to-treat-style impacts on the set of eligible children, many of whom had limited treatment exposure.

Third, we also find positive treatment effects on child nutrition. The intervention increased weight-for-age z-scores (WAZ) by 0.10σ (p<0.01) and height-for-age z-scores (HAZ) by 0.09σ (p<0.05) in AWC endline measurements. Children in treated centers were 3.1 percentage points less likely to be severely malnourished, defined as a WAZ score below -3σ ($p < 0.05$). This represents a 34% reduction relative to the control mean of 9.1%. The treatment also reduced stunting and severe stunting (defined as HAZ scores below -2σ and -3σ), by 4.8 and 2.3 percentage points ($p < 0.05$). These represent a 16% and 40% reduction relative to control means of 29.1% and 5.7% respectively. Nutrition estimates in the household sample point in the same direction but are smaller and statistically insignificant, likely reflecting the inclusion of children with lower intensity of treatment exposure in this sample.

Fourth, we estimate that the intervention was highly cost-effective. Based on literature estimates, we project that the present discounted value of earnings gains expected to result from this intervention's impacts on learning is likely to be roughly 13 times the cost.[2] We also conduct a sensitivity analysis, which suggests that the program would be cost-effective even under conservative assumptions regarding the economic value of test-score gains. Moreover, our projections suggest that the government would fully recover program costs in present value if it captured as little as 8% of the increased earnings as tax revenue. Following the framework of Hendren and Sprung-Keyser (2020), this suggests that the marginal value of public funds invested in the program could be very high since the program is likely to pay for itself over time and generate large additional gains to citizens. In a parallel RCT in the same setting, we found that an unconditional pay increase to existing AWWs had no impact on either education or nutrition outcomes. Thus, the ECE facilitator intervention was highly cost-effective, both in absolute terms and relative to the most common alternative use of funds within the ICDS.

[2] These projections are based on the results in the household sample, which are our preferred estimates for cost-effectiveness calculations since they have a much higher follow up rate. Hence, they do not include the projected benefits of improved nutrition, since these results are not significant in the household sample. Incorporating projected benefits of nutrition gains for the AWC sample increases the estimated benefit-cost ratio to between 17 and 22 (see Section 5).

Our first and most important contribution is to present experimental evidence that it is possible to improve early childhood education with an easily scalable, cost-effective intervention implemented by the government, and to do so in the context of the largest early childhood care program in the world. Expansions in access to pre-primary education in upper-middle income countries have been found to improve pre-primary school attendance and learning (Berlinski, Galiani, and Manacorda, 2008; Berlinski, Galiani, and Gertler, 2009). Yet, public preschool expansions in lower-income countries have been less effective, perhaps reflecting weaker state capacity for implementation (Bouguen et al., 2018; Blimpo et al., 2022). Successful programs in these settings have typically been operated by non-government entities (Martinez, Naudeau, and Pereira, 2017; Dean and Jayachandran, 2019). Our results suggest that strengthening existing public preschool education systems by adding staff can be an effective option for improving early childhood education outcomes at scale.

Second, while there is considerable evidence that interventions in the first 1,000 days of life (including in-utero) can improve child nutrition (see, e.g., Britto et al., 2017), there is much less evidence on whether it is possible to reduce child stunting after this period. We contribute to the child-nutrition literature by presenting experimental evidence that it may be possible for interventions to promote "catch-up" growth among children age 3 to 5.[3] Our results suggest that augmenting front-line staff strength in early childhood programs can be a cost-effective way of improving early childhood nutrition outcomes as well.

Third, we contribute to the literature on building state capacity for service delivery in developing countries. Low-income countries typically have a much lower ratio of public employees per citizen in part because of their lower tax-to-GDP ratios, and in part because of much higher public-employee salaries relative to GDP than richer countries (Finan, Olken, and Pande, 2017). Further, a growing body of evidence suggests that this civil-service wage premium is not correlated with productivity (Bau and Das, 2020; de Ree et al., 2018), and that limited staffing adversely affects service delivery (Dasgupta and Kapur, 2020). Our results suggest that hiring of community-level staff at lower than civil-service salaries may be a promising and cost-effective policy option for expanding state capacity for service delivery more broadly (Haines et al., 2007; Muralidharan, 2016). They also highlight the potentially large economic returns to investing in state capacity in developing countries.

Fourth, our results speak to the literature on the costs and benefits of occupational licensing (Kleiner, 2000). Policy initiatives for expanding early childhood education often stipulate that teachers should be qualified and trained (Berlinski and Schady, 2015; GoI, 2020). Our results, finding that locally hired staff with a secondary-school education and just a week of training

[3]Barham, Macours, and Maluccio (2013) present indirect experimental evidence suggesting the existence of catch-up growth. Gelli et al. (2019) present experimental evidence and Singh, Park, and Dercon (2014) present panel-data evidence that school feeding programs can contribute to catch-up growth between ages 5 and 8. But experimental evidence on catch-up growth is very sparse. See Singh (2014) for a discussion.

were highly effective at improving learning outcomes, suggest that a lack of qualifications may not be a constraint to educator effectiveness in settings with very low student learning levels. These findings are consistent with similar results in the context of primary-school education (Banerjee et al., 2007; Muralidharan and Sundararaman, 2013).[4]

Finally, we provide new evidence on the efficacy of increased instructional resources in education production. While empirical evidence on class-size reductions in low- and middle-income countries is mixed (Urquiola, 2006; Banerjee et al., 2007), smaller class sizes may be especially beneficial for younger children (Blatchford and Mortimore, 1994; Lazear, 2001). Our results support this hypothesis by showing that adding instructional staff can generate large benefits for young children.[5] More generally, our experimental results from India are consistent with and complement historical evidence from higher-income countries that has found large long-term benefits from investing in early childhood education and nutrition programs (Alex-Petersen, Lundborg, and Rooth, 2017; Hendren and Sprung-Keyser, 2020).

2 Setting and intervention

India has over 160 million children between age 0 and 6. Recent data show that 36% of Indian children are stunted and 32% are undernourished (MHFW, 2020). Put together, India has the world's largest number of malnourished children, which significantly increases their risk of not reaching their developmental potential (Lu, Black, and Richter, 2016).

India also faces a severe challenge of low learning levels, with 50% of rural students in fifth grade not able read at a second grade level (ASER, 2019). The challenges start early: the same survey found that 43% of first-graders could not recognize letters and 36% could not recognize one-digit numbers. Learning is particularly poor among public-school students: only 19% of public-school first graders could read words, compared to 42% of those in private schools (ASER, 2020). This likely reflects the greater number of first-generation students in public schools. It also highlights the potential importance of high-quality early childhood interventions to bridge gaps in school readiness and basic skills.

India's national policy documents have long recognized the importance of early childhood education. The 86th Amendment to the Indian Constitution in 2002 directed states to "provide [early childhood education] to all children until they complete the age of 6." The Right to Education Act of 2009 promoted the free and public provision of education for children ages 3

[4]This result may only apply to settings of very low student learning. Evidence from upper-middle income countries suggests that additional teacher qualifications may be needed in settings where most children have mastered basic skills (Andrew et al., 2019).

[5]Class size reductions may also not help much in older grades because they may not alleviate the binding constraint that student learning levels are often several grade-levels behind curricular (and instruction) standards (Banerjee et al., 2007; Duflo, Dupas, and Kremer, 2011; Muralidharan, Singh, and Ganimian, 2019). This concern is less likely to apply to much younger preschool children.

to 6. The National Early Childhood Care and Education Policy Framework, adopted in 2013, stipulated the broad domains of child development that preschool education should cover. Finally, the National Education Policy of 2020 aims that all children ages 3 to 6 should have access to "free, safe, high quality, developmentally appropriate care and education by 2025."

Achieving these policy aspirations has been difficult in part due to constraints in funding and state capacity for implementation (Prasad and Sinha, 2015). More generally, India is characterized by substantial gaps between the aspirations set out in policy documents and the quality of delivery in practice (see, for instance, Pritchett, 2009). Thus, the key challenge for early childhood development in India is not so much at the level of policy intentions, but more so at the level of augmenting capacity for implementation.

2.1 The Integrated Child Development Services (ICDS)

The ICDS is the main public program through which the Government of India promotes early childhood development in India. ICDS provides all of its services through *anganwadi* centers (AWCs). Each AWC serves a catchment area of 400-800 people, and is typically staffed with one *anganwadi* worker (AWW) and one *anganwadi* helper (PEO, 2011).

The worker is responsible for all services provided at the center, spanning early childhood health, nutrition, preschool education, and administrative duties, with duties in both the center and in the broader community. Center-level tasks include early childhood education, overseeing school feeding programs, and providing nutritional supplements. Community-level duties include conducting home visits to raise awareness of appropriate nutritional and health practices; growth monitoring of children; providing supplemental nutrition packets to undernourished children; and coordinating with local nurses to organize immunization camps and health check-ups for children enrolled in AWCs. In addition, the workers have a considerable amount of administrative work and are expected to maintain as many as 14 different paper registers (PEO, 2011). Finally, they are also frequently asked to assist with other government activities, such as surveying, managing electoral booths, and conducting awareness on public schemes in their community. AWWs are typically female, residents of the local village or urban ward, and between 25 and 35 years of age when hired. Their minimum qualification is to have passed a secondary school (10th grade) exam (ICDS, 2017).[6]

The helpers serve as assistants to the workers, and are primarily responsible for cooking and cleaning. Their duties include picking up children from their homes and taking them to the center, cleaning and maintaining the center, teaching children to use the toilet, and helping them to maintain personal hygiene and cleanliness. Helpers are also responsible for preparing,

[6] *Anganwadi* workers receive a monthly honorarium, which is financed by the central and state governments. On October 1, 2018, the central government raised its contribution from INR 3,000 to INR 4,500 per worker per month (AI, 2021). States' top-ups vary widely, from no additional funds (e.g., in Arunachal Pradesh and Nagaland) to over INR 7,000 (e.g., in Haryana and Madhya Pradesh, MWCD, 2019).

cooking, and distributing meals and nutritional supplements. Unlike AWWs, helpers are not subject to formal education requirements beyond the ability to read and write (GoTN, 2021). In our data, less than 40% of helpers had completed middle school (grade 8) and only 11% had completed secondary school (grade 10).

Several non-experimental studies have found positive impacts of ICDS on a wide range of human development outcomes. For instance, Hazarika and Viren (2013) find that children who attend AWCs during ages 0 to 6 are more likely to enroll in primary school; Nandi, Behrman, and Laxminarayan (2020) find that children who attend AWCs in their first three years of life complete more years of schooling; and Ravindran (2020) reports that children who were born in geographic areas with a higher concentration of AWCs were less likely to be underweight and had better early numeracy and literacy skills. Further, early nutritional interventions delivered through ICDS have been found to boost primary-school enrollment, educational attainment, marital age, and employment (Nandi et al., 2016; Nandi et al., 2018).

Advocates for the ICDS and children's rights have frequently called for increasing public spending on the ICDS, including increasing the salaries of existing workers, and hiring an additional worker (see, e.g., Sinha, 2006; Working Group for Children Under Six, 2012; Sinha, Gupta, and Shriyan, 2021). However, despite evidence on the positive impacts of ICDS as a whole, there is much less evidence on the impact of expanding public spending on the ICDS, and on the relative effectiveness of different ways of doing so.[7] Our study contributes experimental evidence of impact to inform this debate, with a focus on children of age 3-6.

2.2 The early childhood education (ECE) facilitator intervention

Our study is set in the southern Indian state of Tamil Nadu, with a population of 68 million people and around 4.2 million children aged 3-6. Child-nutrition outcomes are better than national averages, but still concerning by absolute standards: 25% of children are stunted and 22% are undernourished (MHFW, 2020).

The Government of Tamil Nadu (GoTN) sought our inputs on ideas worth testing to improve outcomes in the ICDS in a cost-effective and scalable way. Since existing research on school education in India had shown that learning deficits appear early (especially for first-generation learners), we identified improving the quality of preschool education as a promising idea to consider. We also conducted a diagnostic study on worker time-use in 24 centers across urban, rural, and tribal districts and found that they spent only 38 minutes per day on preschool instruction on average. Further, in qualitative surveys conducted for the

[7] One recent exception is World Bank (2018a) which experimentally studies the impact of adding a daycare facility to AWCs (in the state of Madhya Pradesh) where working mothers could drop off children under 3. The study found no impacts on either nutrition or education outcomes of children. Since the program was randomized at the community level, the non-impact could also reflect low take up of the program: there was only an 8.2 percentage point increase in the receipt of early childhood services in treated communities.

diagnostic study, workers frequently mentioned that centers were understaffed relative to their responsibilities. We therefore proposed to pilot and evaluate the impact of providing AWCs with an extra staff member to focus on early childhood education.

The intervention we study provided randomly-selected centers the opportunity to hire an extra early childhood education (ECE) facilitator to focus on preschool instructional tasks. Facilitators were hired on two-year contracts using a similar set of eligibility criteria to those used for *anganwadi* workers, though the minimum age was 18 rather than 25 years. They were expected to arrive at the center by 9:45am and provide preschool education from 10am to 12pm. They were expected to work half the hours of workers, and were correspondingly paid around half their salary (Rs. 4000/month compared to Rs. 8000/month).

GoTN had already developed instructional content for ECE (in partnership with UNICEF) and created materials for training AWWs in implementing this curriculum. GoTN developed training manuals for facilitators based on the same materials, and provided them with one week of training. GoTN's communications to field staff noted that the goal of the program was to both improve the quantity and quality of instruction (through the dedicated facilitator) and also to improve child health and nutrition outcomes (through freeing up time of the worker to focus more on these activities). Thus, the intervention did not change the goals of *anganwadi* centers, but augmented their capacity to deliver these goals.

The addition of the facilitator could improve outcomes in several ways, including decreasing the likelihood of centers opening late or being closed; increasing preschool instructional time; enabling instruction in smaller groups if the worker and facilitator teach simultaneously; and increasing workers' time available for health and nutrition related tasks. Our results should thus be interpreted as the composite effect of the intervention through all of these channels.

3 Research methods

Our design and methods follow a registered pre-analysis plan.[8] All analyses in the tables and figures in the main text of the paper were prespecified in this plan.

3.1 Sampling, randomization, and implementation quality

We randomly sampled four districts across the state, to be representative of a population of 60 million people.[9] For ICDS administration, each district is divided into projects comprising 100-150 *anganwadi* centers, which are in turn divided into sectors comprising 15-30 centers

[8] See `https://www.socialscienceregistry.org/trials/1772`.

[9] We excluded the district of Chennai, which is the state capital and a metropolis of over 7 million people. District sampling was stratified by geographic zones and average nutrition status (see Figure A.1). Table B.1 shows that the four sampled districts are very similar to non-sampled districts. For inference, we condition on the set of sampled districts, with standard errors and confidence intervals designed to reflect uncertainty about

each (PEO, 2011). We started with the universe of AWCs in the four sampled districts and excluded those with other NGO interventions, in buildings shared with other centers, and with vacancies in both staff positions (worker and helper).[10] We then randomly sampled 320 centers from the remaining population, stratifying by staffing vacancy and project.

We randomly assigned centers to the control or treatment groups, stratifying randomization by district, an indicator for whether a center had a vacant AWW position, and a principal component of local demographic characteristics.[11] We divided our sample into 40 strata defined by district, vacancy status, and quintiles of the principal component. Within each stratum, we assigned four centers to the control group and four to the treatment group, for a total of 160 control and 160 treatment centers.[12]

Table 1 presents summary statistics on centers, workers, and children, and also compares these baseline characteristics across treatment and control centers. AWCs on average had 15 children enrolled across all ages. Workers on average were around 50 years old, had over 20 years of experience, and were paid around 8,000 Indian Rupees (INR) per month. Around 86% of them had completed secondary school (grade 10) or more.[13] Children attending the centers were 3.5 years old on average. Baseline nutrition levels were low: 37% of children were underweight with a weight-for-age z-score (WAZ-score) below -2σ, and 35% were stunted with a height-for-age z-score (HAZ-score) below -2σ. Consistent with AWCs enrolling children from relatively disadvantaged families, the fraction of underweight children in our sample is higher than the state-wide rate of 24% in rural Tamil Nadu (MHFW, 2020).

We find no systematic differences between the treatment and control groups in center or student characteristics, including on baseline math, language, and executive-function test scores. By chance, workers in control centers were slightly older and more experienced, but slightly less likely to have completed secondary schooling. We control for baseline test scores, randomization stratum fixed effects, and AWW education and experience (to account for the small imbalances we see at baseline) in our main estimating equations.

population parameters for centers in these four districts. Given our sampling process, conducting inference on the full rural population would require clustering at the district level, which is not feasible with four districts.

[10] These restrictions excluded 10.8% of centers from the sampling frame.

[11] These included population, age distribution, language, occupation distribution, and family income based on administrative data for each AWC catchment area.

[12] This project was carried out as part of an institutional partnership between J-PAL South Asia and the Government of Tamil Nadu, under which we studied multiple interventions to improve early childhood education and nutrition outcomes. The three other interventions studied included an unconditional increase in AWW pay, a performance-based bonus to workers based on improvements in child nutrition, and a supplemental feeding program. Results from these interventions are reported in a companion report (Ganimian, Muralidharan, and Walters, 2020). No center received more than one treatment. Thus, there are no interactions across treatments, allowing our estimates to be interpreted as effects relative to a "business as usual" counterfactual (Muralidharan, Romero, and Wüthrich, 2021).

[13] While *current* recruitment norms for both workers and facilitators require secondary school completion, the existing stock of workers includes those hired many years earlier under lower minimum education norms.

GoTN implemented the intervention well. In our first process monitoring survey, conducted five months after GoTN issued the program notification, 98% of treated centers had a facilitator (Table A.1). On average, they had been hired 135 days prior to the survey, confirming that they were hired promptly, within 15-30 days of the notification. Further, nearly all facilitators (96%) reported having received the six-day training required by GoTN.

3.2 Data and attrition

Our core study sample consists of children present in study centers at baseline. Our primary outcomes of interest are these children's scores on independent tests of math, language, and executive-function skills.[14] Tests were administered individually by J-PAL enumerators in a baseline round prior to randomization (September-November 2016) as well as in an endline round 16 months after program rollout (March-April 2018). The test instruments were designed to minimize ceiling and floor effects and produce a distribution with broad support. Baseline test scores are standardized ($\mu = 0$, $\sigma = 1$) in the full sample, and endline scores are standardized relative to the control group distribution. Appendix C provides more details on test construction, characteristics, and administration.

As per our pre-analysis plan, we also study treatment effects on child nutrition as a secondary outcome. Our main measures of nutrition are WAZ and HAZ scores. We study impacts on average WAZ and HAZ as well as on proportions of children with scores below -2σ and -3σ, which are widely used measures of moderate and severe malnutrition and stunting. Since measurement of child anthropometric data can be sensitive to field protocols, enumerators received extensive training, and each child was measured twice. Appendix C provides further details on observation, measurement, and training protocols.

At endline, enumerators visited every center twice within a week to measure outcomes for as many children as possible. There was no difference in follow-up rates across treatment and control groups (Table 2, Panel A, col. 1). We also see no differences in composition across treatment and control groups along student age, gender, baseline test score, or nutrition status, with all interactions of treatment status and these characteristics insignificant in a model for follow-up (Panel B, col. 1). A joint test of significance across all interactions confirms that there was no differential attrition between treatment and control groups across observed baseline characteristics, though this does not guarantee balance on unobservables.

However, the overall follow-up rate for children in the baseline AWC sample was only 33%. This likely reflects a combination of children moving out of the *anganwadi* centers to private preschools, attending irregularly, migrating, and ageing out of preschool and enrolling in first

[14] We measured executive function by assessing children's inhibitory control, working memory, and cognitive flexibility. These provide measures of cognitive development that are independent of curricular content.

grade at age 5 (Table 2 shows that older children are significantly more likely to attrit).[15] Since there was no differential attrition, we interpret treatment effects in the AWC sample as representing program effects for children who stayed enrolled in the centers during the study and were likely to attend regularly.

To estimate intent-to-treat (ITT) effects on the entire baseline sample (including those who did not stay enrolled in or attend regularly), we supplement the AWC-based measurements with household-based measurements. Specifically, we drew a representative sample of 50% of children who were observed in the baseline sample, visited their households, and tested them there in May-June 2018 (18 months after program rollout).[16] There was again no differential attrition between treatment and control groups overall or by observable baseline characteristics (Table 2, col. 2). However, the follow-up rate for the household measurements was much higher, at 89% compared to 33% for the AWC endline.

We also conducted one round of unannounced and announced visits to centers over the course of the study. The unannounced visits were used to measure attendance, punctuality, and time use. Enumerators arrived at each center before the official opening time to determine when the center opened, and when the worker and the facilitator arrived. They then tracked the amount of time that the worker and facilitator spent on various tasks between 10am-12pm (the scheduled time for preschool instruction), using an adaptation of the Stallings Observation System (see Stallings and Mohlman, 1990). We collected data based on these observations in a random sample of 40 centers per district (20 each in the treatment and control groups), for a total of 160 centers (50% of the study universe of 320 centers). The announced visits were used to survey workers and facilitators and to obtain additional details on teaching practices.

3.3 Estimation

Our main equation for estimating program impacts is:

$$Y_{ic} = \alpha_{s(c)} + X'_{ic}\gamma + \beta T_c + \varepsilon_{ic}, \tag{1}$$

where Y_{ic} is an outcome for child i enrolled at center c; $s(c)$ is the randomization stratum of center c and $\alpha_{s(c)}$ is a stratum fixed effect; X_{ic} is a vector of baseline covariates that includes a baseline measure of the outcome variable for individual children, the mean baseline outcome for all children at the center, and AWW education and experience; T_c is an indicator equal to one if center c is assigned to the treatment group; and ε_{ic} is an error term.

[15] Household survey data from a different study in the same districts (Singh, Romero, and Muralidharan, 2022) also shows that the share of children enrolled in AWCs drops sharply from age 3 to age 5 (Figure A.2.)

[16] To increase precision for studying effects on malnourishment, we oversampled children with WAZ$<$-2σ. All results using this household sample are reweighted to be representative of the full baseline sample. Since the household survey sample was drawn from the set of children initially observed at the AWC, neither sample includes children who were absent from *anganwadi* centers at baseline.

The parameter of interest is β, which represents the average causal effect of a center receiving the ECE facilitator intervention. We estimate equation (1) by OLS regression in the AWC sample. Regressions in the household sample are weighted to account for differences in sampling probabilities, allowing us to recover effects for the population of children who took the baseline test. Standard errors are clustered at the AWC level.

4 Results

4.1 Center openings and staff attendance

The addition of the ECE facilitator significantly reduced the likelihood of centers being closed at the scheduled start of preschool instruction (10am). Based on our unannounced visits, treated centers were closed only 3.8% of the time compared to 12.5% in control centers, which is a 70% reduction (Table 3, Panel A).[17] Centers were also around 5 percentage points more likely to be open by the scheduled opening time of 9am (though the difference is not significant). Overall, the presence of an extra staff member improved center quality on the extensive margin of the likelihood of centers being open and opening on time.

The addition of the facilitator also reduced the absence of *anganwadi* workers in treated centers by 50%, from 20% to 10% (Table 3, Panel B). This may reflect the need for workers to arrive in time to open the center for the facilitator. Since AWWs are responsible for items stored in the center (including provisions for feeding children), centers are typically locked with the worker having the keys. Consistent with this idea, the absence rate of workers in treated centers (10%) was similar to that of the facilitators (8.7%), whereas it was significantly higher in control centers (20%). This complementarity between worker and facilitator attendance may have contributed to increased worker attendance in treated centers.[18]

4.2 Time use

Next, we examine impacts on the intensive margin of time use during the two-hour window of direct observation of classroom activity during the time scheduled for preschool instruction (10am-12pm). We find that facilitators spent around half this time (57 minutes) on preschool instruction (Table 4, col. 1). They spent around 20 minutes on administrative work, and

[17] The estimates in column 3 differ slightly from the gap between columns 1 and 2 because column 3 controls for worker characteristics. These regressions exclude randomization strata controls because the random sample of visits was not stratified, so some strata include zero visited centers. Note that controls for strata are not necessary for unbiased treatment effect estimation because the probability of treatment is equal across strata.

[18] This finding contrasts with that of Muralidharan and Sundararaman (2013) and Duflo, Dupas, and Kremer (2015) who find that adding a contract teacher to schools in India and Kenya *reduced* attendance of the existing teachers.

6 minutes on health and nutrition tasks. The remaining 37 minutes were either off-task (27 minutes) or accounted for by absence (10 minutes).[19]

The intervention also shifted the time allocation of *angwanwadi* workers in the expected direction. Worker time spent on preschool education fell roughly in half in treated centers: the average AWW in the control group spent 38 minutes per day teaching (col. 2), whereas her treated counterpart spent only 18 minutes per day (col. 3). However, workers in treated centers increased time spent on administrative tasks such as completing paperwork (35 v. 22 minutes), and on health and nutrition tasks (11 v. 6 minutes).[20] All three differences above are significant at the 1% level (col. 4). Workers in treated centers increased their time off duty (uninvolved, out of the center, or engaged in social interactions), but this was offset by a corresponding reduction in time off task due to absence. This suggests that the presence of the facilitator may have increased shirking among AWWs while on the job, though the offsetting decline in time lost to absences led to total AWW time spent on education, administrative tasks, or health and nutrition tasks being roughly unchanged.[21]

Despite the reduction in instructional time by AWWs, the intervention led to a large increase in total time spent on early childhood education. Adding the time spent by both the worker and facilitator, children in treated centers received 76 minutes per day of preschool instruction (Table 4, col. 5), effectively *doubling* the time allotted to education relative to the control mean of 38 minutes (col. 6 vs. col. 2). The intervention also led to a near-tripling of time spent on health and nutrition related tasks (an increase of 11.3 minutes from a control mean of 5.7 minutes), and about 2.5 times more time spent on administrative work. Overall, the addition of the facilitator led to an increase in total staff time spent on all major activities, including preschool education, health and nutrition, and administrative work (col. 6).

Since the time window we observe was the part of the day scheduled for preschool instruction, it is unsurprising that the biggest absolute impact on time spent was on preschool education. However, two additional considerations suggest that total time spent on health and nutrition activities may have increased more than the impacts we measure in our two-hour observation window. First, data on self-reported time use from the facilitators suggests that they spent around 1.5 hours per week on health-related activities (Table B.2), substantially more than the six minutes per day we see in the observation window. Second, the

[19]When a staff member is absent, we code the entire two-hour observation window as absent. A similar approach is used for partial attendance. So, if a staff member arrived at 10:30am, they would be coded as absent for the first 30 minutes, and their actual activity would be coded for the remaining 90 minutes of the observation window. Thus, the full 120-minute observation window is accounted for in the coding.

[20]The health and nutrition category captures time spent preparing or serving food, assisting children to use the toilet or wash their hands, and miscellaneous health-related activities (see Table A.2 for details).

[21]It is possible that the time we code as workers or facilitators engaging in "social interactions" includes time spent with parents visiting the center. This could be considered as time spent productively if they provided parents with inputs and advice on home feeding practices and interactions with children. We are not able to quantify this since enumerators did not code the identity of who the staff were interacting with.

administrative work for the AWW did not change due to the intervention. Thus, completing some of this work during the time scheduled for ECE (as seen in Panel A), while supervising the instructional work of the facilitator is likely to have freed up AWW time outside the observation window to focus on nutrition and education related activities.

4.3 Learning outcomes

Consistent with the doubling of total time spent on preschool instruction, the provision of an ECE facilitator produced large test-score gains. Children in treated AWCs scored 0.29σ, 0.46σ, and 0.18σ higher in math, language, and executive function on independent tests administered at the AWC 18 months after baseline, with all results being significant at the 1% level (Table 5, Panel A, row 1). On a composite measure of learning constructed as the first principal component across the three tests, children in treated centers scored 0.29σ higher ($p < 0.01$). We also see large gains in learning levels in treated AWCs when outcomes are defined as the proportion of test items answered correctly rather than standardized scores (Table B.3).[22]

Treatment effects on the household tests were smaller but still significant. Children in treated centers scored 0.13σ higher in math ($p < 0.01$), 0.10σ higher in language ($p < 0.05$), 0.05σ higher in executive function (not significant), and 0.11σ higher on the composite score ($p < 0.05$) in the household assessments (Table 5; row 2).

The smaller treatment effects in the household assessments likely reflect the fact that this sample includes children who were no longer attending the AWC by endline. Figure A.3 plots the age distribution of children in the baseline and in both the AWC and household follow-up samples, and clearly shows that the household sample includes many more children over 5, who are likely to have aged out of the AWC. Thus, the lower estimated effects in the household sample likely reflect the inclusion of children with low program exposure. We directly examine this hypothesis by reporting treatment effects on the *common sample* of children who were present for both the AWC and household endline tests. As shown in Panel B of Table 5, treatment effects on the household tests in the common sample are considerably larger than corresponding effects for the full household sample, and very similar to treatment effects in the AWC sample (0.32σ vs. 0.29σ for composite scores).[23]

Since there is no differential attrition between treatment and control groups in either sample, we interpret the larger effects found in AWC sample as reflecting the experience of children who continued to actively attend *anganwadi* centers during the study, and the smaller

[22] Table B.4 reports impacts on the likelihood that a student answers *any* items correctly. We see small positive effects on this outcome on HH assessments and insignificant estimates for AWC assessments, suggesting that the large impacts observed at the AWC are primarily on the intensive margin.

[23] As described in Appendix C, the household assessment was shorter and included a subset of items from the AWC assessment. Table B.5 shows that we also find similar effects on household and AWC assessments for the common sample when the analysis is limited to the common test items.

impacts in the household sample as ITT effects for children present in the AWC at baseline.[24] If we assume that all treatment effects in the household sample accrue to children in the AWC sample, we can obtain an estimate of the effect of treatment on the treated (TOT) by dividing the household sample estimate by the share of household sample children with endline AWC scores. This produces a composite score TOT of $\frac{0.11\sigma}{791/2080} = 0.29\sigma$, an estimate that is very close to the corresponding effects for the common sample and complete AWC sample ($0.29 - 0.32\sigma$). This is consistent with the assumption that the difference between household and AWC estimates reflects more intense treatment exposure for the AWC sample, though some children in the household sample who were not captured in the AWC assessments may have received some benefits from the intervention.[25]

Finally, we find limited evidence of heterogeneity in treatment effects. Quantile treatment effect plots for the AWC assessments show that the treatment distribution first-order stochastically dominates the control distribution, suggesting broad-based test-score gains from the program (Figure A.5, Panel A). Non-parametric estimates of average treatment effects at each percentile of baseline composite score show large positive impacts across the full range of baseline achievement (Figure A.5, Panel B). We also find no evidence of differential effects by baseline nutrition (weight-for-age z-score), mothers' education, or AWW vacancy (Table A.4). We find suggestive evidence of greater effects on girls, but this result may reflect chance variation as it is not significant in the common sample.

4.4 Nutrition outcomes

We estimate that the intervention also improved nutrition in the AWC sample. Table 6 shows that average WAZ scores were 0.10σ higher in the treatment group ($p < 0.01$). We find no significant reduction in the probability that children are underweight (col. 2), but a significant 3.1 percentage point decline in the rate of severe malnutrition (col. 3; $p < 0.05$). This represents a 34% reduction in severe malnutrition relative to the control mean of 9.1%.

[24] Even with no differential attrition between treatment and control groups and no differences in attrition according to observed characteristics (as shown in Table 2), selective attrition may lead results for the endline sample to be unrepresentative of the full baseline population. To probe robustness to this issue, Appendix Table A.3 reports inverse probability weighted (IPW) estimates based on probit models for follow-up fit separately for the treatment and control groups as a function of baseline test scores, gender, and randomization strata. The IPW estimates are very similar to the baseline estimates, suggesting that selective attrition on these observed dimensions has little effect on the results. As a second robustness check, Appendix Figure A.4 shows a sensitivity analysis using the strategy of Kling and Liebman (2004), which imputes missing outcomes for attriting children. Missing outcomes for control children are imputed as the control mean plus the control standard deviation times a factor Δ, while those for treated children are imputed as the treated mean minus the treated standard deviation times Δ. The impact estimate for composite scores in the AWC sample remains positive and statistically significant for values of Δ less than 0.2, indicating that outcomes for attriting children would have to differ substantially from the observed sample to reverse the conclusions of our analysis.

[25] Similarly, Table A.7 shows that treatment effects in the household assessment are positive and significant for those who self-report attending the AWC, and insignificant for those who report that they do not attend.

Similar patterns are observed for HAZ scores, which were 0.09σ higher in treated AWCs (Table 6, Panel B). The treated group also saw a 16% reduction in the fraction of stunted children (a 4.8 percentage point reduction from a base of 29.1%) and a 42% reduction in the rate of severe stunting (a 2.3 percentage point reduction on a base of 5.7%).

As with the test-score results, the impacts on nutrition and stunting in the household sample are in the same direction as the AWC results, but the magnitudes are smaller, and in this case statistically insignificant. Restricting attention to the common sample, we find similar increases in average WAZ and HAZ scores, and a similar reduction in moderate and severe malnutrition and moderate and severe stunting in the AWC and household measurements; and cannot rule out that the effects are the same in both samples (Table A.5).[26]

We investigate distributional treatment effects on WAZ and HAZ scores in the AWC measurements and find broad-based evidence of positive impacts, with little systematic evidence of effect heterogeneity (see Figure A.6 and Table A.6 for WAZ, and Figure A.7 and Table A.9 for HAZ). We also verify that the nutrition results are not sensitive to measurement outliers. Tables B.6-B.9 show that the estimated WAZ and HAZ impacts are robust to dropping outlier measurements and winsorizing the outcome variables.

These results suggest that the benefits of providing an extra worker to focus on educational activities were not restricted to improving education outcomes, but extended to improving nutrition outcomes as well. This is consistent with the time use data showing increased time spent on health and nutrition related activities.[27] However, as in the case of the test score results, these positive effects are mainly seen in the sample of children who stay enrolled in the AWC and were likely to have attended more regularly, over the 18 months of the study.[28]

Finally, we consider the extent to which improved nutrition may have contributed to the learning gains we find. In the baseline data, the coefficient from a regression of composite test scores on WAZ scores is 0.15, while the corresponding coefficient for HAZ scores is 0.14 (Table A.10). These correlations plausibly represent upper bounds on the causal effects of improved nutrition since omitted variables correlated with nutrition and learning are likely to affect both outcomes in the same direction. This logic suggests that the 0.10σ improvement in WAZ scores in the AWC sample contributed at most $0.1\sigma \times 0.15 = 0.015\sigma$ of the improvement in composite test scores, representing around 5% of the total test score impact (0.29σ). Thus, the direct mechanism of extra instructional time enabled by the ECE facilitator likely accounts for the majority of the test-score gains we see, with the improved nutrition outcomes enabled by freeing up time of the incumbent worker likely being a second-order channel.

[26]IPW estimates adjusting for attrition on observables are very similar to unweighted results (Table A.8).

[27]As noted in Section 4.2, total staff time spent on health and nutrition tasks could have also increased further outside the observation window. For instance, having the facilitator may have helped the AWW to ensure that malnourished children got adequately fed during the lunch provided at the AWCs at 12pm.

[28]Similar to the test-score results, Tables A.11 and A.12 show more positive WAZ and HAZ estimates in the household sample for those who self-report AWC attendance.

5 Cost-effectiveness

We analyze the cost-effectiveness of the ECE facilitator intervention in two ways. First, we assign an economic value to the program by calculating the present discounted value (PDV) of projected improvements in participants' future earnings from the estimated short-run treatment effects. Comparing this value to program costs yields a benefit-cost ratio and an estimate of the rate of return from investing in the program. From a public finance perspective, this is the estimate that matters since it informs the marginal allocation of public funds. However, the limitation of this approach is that it relies on strong assumptions to project future earnings gains. We therefore also compare the cost effectiveness of the program relative to alternative uses of funds *within* the ICDS.

Our benefit-cost calculation focuses on the program's test-score impact for the household sample, which represents our best estimate of the intervention's average impact on the full population of children present at baseline. This calculation ignores any nutrition benefits because the impacts on nutrition are not statistically significant in the household sample.

Since we will not be able to measure labor-market outcomes for children in our sample for many years, we use global-literature estimates of the relationships between impacts on short-run and long-run outcomes collected in Kline and Walters (2016) to project the impacts of the intervention on future earnings. As emphasized by Heckman et al. (2021), such projections may understate the long-term benefits of early childhood programs if their impacts operate through non-cognitive channels. Our calculations are also conservative in ignoring non-labor market benefits of better health and education. These include the intrinsic value of better health and education for citizen well-being (Sen, 2001), as well as the instrumental benefits of improved education on better decision making in areas ranging from health behaviors to personal finance (see, e.g., Vogl, 2012; Cole, Paulson, and Shastry, 2016).

The cost-effectiveness analysis is presented in Table 7. Panel A lists the assumptions used to calculate the expected PDV of future earnings for the control group. Predicted labor-force participation rates and daily wages are measured based on statistics for the rural Tamil Nadu population from the 2011-2012 National Sample Survey (NSS). We assume that workers are employed for 225 days per year when in the labor force and that wages will grow at a real annual rate of 5%, which is conservative compared to the 6-7% growth rate of real GDP per capita in Tamil Nadu during 2012-2019.[29] We assume people work from age 22 to 65 and discount these projections back to age 4 using a discount rate of 3 percent.[30] These parameter

[29]The NSS reports that on average, people report working a little over 6 days per week throughout the year, which would imply 300 working days a year (see Figure 1D in Muralidharan, Niehaus, and Sukhtankar, 2021). However, this includes self-employment, which may have lower marginal product than paid market labor. We therefore use a more conservative assumption of 225 working days per year.

[30]The working age assumption is likely to be conservative since many children in Tamil Nadu may start working at earlier ages. The discount rate is based on the state government's real cost of borrowing, which

values imply that the average PDV of total future earnings for children in rural Tamil Nadu equals roughly INR 3.6 million.

Panel B combines this PDV projection with the experimental treatment effects to predict the present value of earnings gains from the intervention. This exercise requires an assumption linking the program's short-term impacts to its long-term effects on earnings. Kline and Walters (2016, Appendix Table A.IV) document that the ratio of percentage earnings gains to standard deviations of test score gains is 10 percent or more for a variety of educational interventions in disparate settings. A key benchmark comes from Chetty et al. (2011), who report a ratio of 13 percent in a long-term follow-up of an experimental study of kindergarten class quality. This estimate may overstate the value of test score gains in Tamil Nadu if many children work in agriculture or the informal labor market where the returns to academic achievement may be low. At the same time, other evidence suggests that early education interventions may increase educational attainment and other long-term outcomes while having limited or short-lived impacts on test scores (Garces, Thomas, and Currie, 2002; Bailey, Timpe, and Sun, 2021; Chetty et al., 2011; Chetty, Friedman, and Rockoff, 2014), in which case projecting the program's value based on its test score effects may understate its benefits.

Using this estimate of an earnings impact of 13% per standard deviation of test-score gains, our experimental treatment effect of 0.11σ for the household sample (Table 5, Panel A, col. 4) yields a projected earnings increase of $0.13 \times 0.11 \times 100 = 1.4$ percent. Applied to the PDV reported in Panel A, we project that the intervention will boost the PDV of future earnings by about INR 52,000 per child. Our experimental sample included 14 children per treated center at baseline. Since the experiment was conducted over 18 months and children typically age out of the center after two years, the yearly count of 14 is roughly three-fourths of the total number of children treated by the program. This is because our analysis excludes the new cohort of children who would have joined the center in the second year (for whom we do not have a baseline test). We therefore multiply the projected earnings benefit per child by 14×1.33 to calculate a total expected gain from the program, which equals about INR 964,000. The cost of the program was roughly INR 74,000 per center over eighteen months.[31] Taking the ratio of these two numbers produces a predicted benefit-cost ratio of 12.9.[32] The details of this calculation appear in Appendix D.

Figure 1 assesses the sensitivity of the estimated benefit-cost ratio by varying each of the calibrated parameters in Table 7. Specifically, we draw each unknown parameter from

is 2-3%. Yields on 10-year government bonds during our experiment were around 6-7%, whereas the inflation rate was 4-5%.

[31] This cost includes a one-time training cost for each facilitator along with 18 months of salary and administrative costs for the program.

[32] Thus, even if the government were to hire a second AWW at the regular pay of an AWW (which is double that of the facilitator), such an investment would likely be cost-effective. Of course, a regular worker would work a full shift and the benefits may be correspondingly larger.

a distribution defined on a range of possible values with the values from Table 7 in the center. Panel A draws the parameters determining days worked in the labor market, wage growth, the discount rate, and the earnings premium for test scores independently from uniform distributions, and panel B draws them from truncated normal distributions. The benefit-cost ratio is large for essentially all parameter values we consider. For Panel A with uniform parameters, the 5-95 percentile range of benefit-cost ratio is 4.2 to 39.1. For Panel B, the range is 5.5 to 39.0. Thus, even under very conservative realizations of the parameters used to value test score gains, the benefits are likely to be greater than the costs.[33]

Our preferred (and more conservative) estimate of benefit-cost ratio is the one above using estimates from the household sample. For completeness, we also present an alternative calculation using estimates based on the AWC sample in Table A.13. The AWC sample produces larger test-score gains for a smaller number of children, resulting in a very similar benefit-cost ratio of 12.2. However, since we find significant nutrition gains in the AWC sample, we also consider a projection adding earnings gains from improved nutrition. Doing so requires additional assumptions to project the economic value of nutrition gains and to combine impacts through multiple channels, since some of the program's learning gains may be caused by improved nutrition. We project the earnings impact of improved nutrition based on Hoddinott et al. (2011) and Hoddinott et al. (2013)'s analysis of the labor-market impacts and economic value of improved childhood nutrition in Guatemala. To avoid double-counting the program's nutrition effects, we subtract the estimated improvement in test scores attributable to nutrition based on the cross-sectional correlation between nutrition and test scores before assigning a value to the test score gains. Including nutrition effects measured by improvements in HAZ (col. 3) or reductions in stunting (col. 4) boosts the benefit-cost ratio to between 17 and 22 (Table A.13, columns 3-4).

Finally, from a public finance perspective, we also need to account for the expected increase in future tax revenue from increasing the earnings of citizens. Hendren and Sprung-Keyser (2020) suggest prioritizing government programs based on the marginal value of public funds (MVPF), defined as after-tax benefits to participants per dollar of net cost to government inclusive of any impacts on tax revenue. The projections in Table 7 indicate that as long as children pay a net tax rate of at least $(74,478/964,439) \times 100 = 7.7\%$ on future earnings, program costs are more than paid back in present value by expected future tax revenues.[34] In a historical analysis of returns to social programs in the United States,

[33]Cost/benefit calculations may be especially sensitive to the discount rate. We assess this via alternative simulations that draw the discount rate from a log-normal distribution with median 3 percent and standard deviation 3 percent. The log-normal specification creates skew and results in more large positive draws of the discount rate. As shown in Figure A.8, this approach yields similar distributions of the benefit/cost ratio, though the lowest quantiles are below those in Figure 1, with 5th percentiles of 1.2 and 1.3 in the two panels.

[34]Even if we assume that none of the children who benefit from the program will ever pay income taxes (since less than 10% of Indians pay income tax), India's indirect Goods and Services Tax (GST) covers most

Hendren and Sprung-Keyser (2020) report that early childhood interventions are among the most cost-effective and frequently generate increases in future tax revenue sufficient to recover the government's costs. Our results suggest that such returns may also be possible for contemporary large-scale early childhood interventions in India.

The benefit-cost calculations presented above are the relevant metric from a public finance perspective.[35] However, one limitation of this approach is that it requires strong assumptions to extrapolate from a program's short-run treatment effects to predict impacts on future economic outcomes. Thus, a second approach is to compare the cost-effectiveness of alternative approaches to improve short-term outcomes. While this does not answer the public finance question of whether to expand funding to the program, it helps answer the question of how budgets *within* a sector or program may be better allocated towards more cost-effective interventions. This is the approach taken in the case of education by institutions such as JPAL and the World Bank who produce evidence syntheses where they report the "cost per standard deviation of test-score gains" across interventions (see, e.g., World Bank, 2021).

Since there is considerable contextual variation in early childhood programs around the world, we limit this approach to comparisons across alternative uses of funds within the ICDS itself. One natural comparison is with the impact of increasing the salaries of existing *anganwadi* workers. Worker salaries are the largest source of ICDS expenditure, and policymakers face pressure from both worker unions and ICDS advocates to increase their pay (see, e.g., The Times of India, 2021). Consistent with this pressure, ICDS budget increases over the past decades have been predominantly used for pay increases for incumbent staff as opposed to hiring more staff. In a companion paper (Ganimian, Muralidharan, and Walters, 2020), we report the impacts of an unconditional across-the-board pay increase to AWWs implemented in another set of AWCs randomly drawn from the same population studied here. This analysis found no significant effects of the across-the-board pay increase intervention on test scores (with negative point estimates), and very limited (and inconsistent) evidence of improvement in nutrition outcomes.

Thus, even without assumptions regarding the mapping from short-term to long-term benefits, our results suggest that early childhood education and nutrition outcomes may be significantly improved by using annual increases in the ICDS budget to hire extra staff rather than increasing the pay of incumbent workers. Overall, our analysis suggests that the ECE facilitator intervention was highly cost-effective, both in absolute terms and relative to the most common alternative use of funds within ICDS.

---

of the economy. Since the GST rate for most commodities is between 12% and 18% it is likely that the government will be able to capture at least 7.7% of increases in expenditure as tax revenue.

[35] In particular, a benefit-cost ratio of 12.9 implies a return on investment of 1190% and an MVPF of infinity at an effective tax rate above 7.7%. Thus, even if governments are fiscally constrained by current tax revenues, the program's expected rate of return would far exceed the cost of borrowing even if program expansion has to be financed by increasing long-term debt.

6 Conclusion

Improving early childhood nutrition and learning outcomes is widely believed to be one of the most valuable investments a country can make in the future of its citizens (Heckman, 2012; World Bank, 2018b). Yet, despite broad agreement on its importance among both academic experts and in global and national policy documents, there is limited evidence to inform *how* low and middle-income countries can achieve this goal at scale. Further, while countries like India have invested in setting up a nationwide program like the ICDS to deliver early childhood services at scale, there is little evidence to inform whether it would be a good use of scarce public funds to augment spending on the ICDS, and if so, on what kind of intervention.

We present experimental evidence on a simple and scalable policy to strengthen the ICDS: hiring an extra staff member to focus on preschool education, and thereby also freeing up time of the existing worker to focus more on child health and nutrition. We find that doing so led to a significant increase in total preschool instructional time and also in time spent on health and nutrition related activities. Consistent with these increases in "time on task," we find that 18 months after the baseline, children who remained enrolled in treated centers had significantly higher test scores in math, language, and executive function. They also had higher WAZ and HAZ scores and lower rates of stunting and malnutrition. From a policy perspective, a key result is that the program was not only effective, but also highly cost-effective in both absolute terms and relative to alternative uses of funds within the ICDS.

There are several features of our study that suggest that the effects we find may be replicated at scale – at least within the state of Tamil Nadu. As noted by List, Suskind, and Supplee (2021), reasons for why impressive gains found in smaller-scale studies may not replicate at larger scales include: (a) poor quality of the original evidence itself, (b) studies conducted in non-representative convenience samples where effects may be larger in sites/individuals who opt in to a study, (c) non-representative 'situations' - for instance, where the interventions evaluated are implemented by motivated non-profit organizations as opposed to typical government bureaucrats, and (d) supply constraints - for instance, it may be easier to hire one good extra teacher than several thousand (at which point supply quality may decline). Our study addresses each of these challenges.

On the first point, the study features a randomized experiment with extensive primary data collection on both processes and outcomes. The second and third concerns are addressed by conducting the study in a sample that is representative of 60 million people of Tamil Nadu (with no center opting out) and by studying a program implemented by the Government itself, using the same protocols that would be used at a larger scale. On the final point, treated centers were able to successfully hire an extra facilitator within 15-30 days of being allowed to do so, with no additional support beyond financial authorization. Further, the facilitator was hired from the same village as the AWC and each village typically has only one AWC.

Since scaling up the program will only require hiring *one* more facilitator in each village, staff supply is unlikely to be a constraint to scaling up.[36]

Additional evidence on scalability is seen from recent experience in the same context where GoTN was able to hire over 200,000 local volunteers (almost all women) to provide supplemental instruction to aid with recovery of COVID-19 learning loss in early 2022. The volunteers were paid a stipend of just Rs. 1000/month to work for 90 minutes a day. This is a lower wage than in our study (which paid Rs. 4,000/month for 240 minutes/day), but the program still attracted nearly four applicants per opening. This program was highly effective as well as cost-effective at improving test scores, and also improved equity (Singh, Romero, and Muralidharan, 2022). These results provide further evidence that the intervention we study may also be effective at scale, and also promote equity (since children from disadvantaged backgrounds are more likely to use the ICDS). It also suggests that the approach we study of using locally-hired staff for augmenting state capacity for child development may be relevant not only for policy decisions on the intensive margin of how to expand existing programs like the ICDS, but potentially on the extensive margin of new programs as well.

Scaling up a two-worker model in the ICDS may have additional benefits for women's empowerment and education beyond the direct benefits on early childhood education and nutrition. Several studies have noted that constraints on traveling outside their village are an important barrier for labor-force participation for young women in rural South Asia (e.g. Andrabi, Das, and Khwaja, 2013). Thus, augmenting ICDS staffing by providing jobs for women in the same village (as done in this study), may also increase female labor force participation, which is especially low in India (Moore, Fletcher, and Pande, 2018). In turn, the expansion of *visible* job opportunities for women may also increase both the true and perceived returns to education for girls in rural areas, resulting in increasing demand for female education over time (Jensen, 2012).

One important caveat to external validity is that the ICDS is believed to work better in Tamil Nadu than in other states with poorer child human development indicators, and that GoTN may therefore have implemented the program better than other states may be able to.[37] Thus, while our study provides strong suggestive evidence on the value of a second worker in the ICDS, further evidence is needed to determine whether the benefits are similar in other states. One promising approach to scaling the intervention we study may be to (a)

[36]The easy availability of facilitators likely reflects the expansion of female secondary-school education across India over the last two decades, and the corresponding increase in supply of secondary-school graduates.

[37]Tamil Nadu had in fact implemented a two-worker model in the ICDS in the 1980s under the Tamil Nadu Integrated Nutrition Project funded by the World Bank and other donors. Qualitative evaluations and interviews suggested that the program was well received at the field level and effective (see Heaver, 2002). However, the second worker was discontinued when donor funding for the project ended. This prior experience partly contributed to GoTN's interest in a high-quality evaluation of the impact of adding a second worker to the ICDS, and to see if it made sense to fund a scale up out of their own budget. The institutional memory of implementing a two-worker model in the ICDS may have also helped GoTN implement the program well.

design a program that combines an additional ICDS staff member while also re-optimizing time and task allocation to include other evidence-based ideas to improve early childhood development,[38] (b) rolling this out at even larger scales (covering say 10,000 - 50,0000 AWCs) across multiple states, and (c) conducting an experimental evaluation using a randomized staggered roll out to evaluate impacts at scale over a longer time horizon.[39] Such an iterative approach could both generate even larger gains than the ones we find (by combining additional staff with the best evidence-based interventions for early childhood development), and would also ensure that decisions on nationwide scale up are informed by evidence from a wider set of contexts, and at even more policy-relevant scales.

References

AI (2021). "Integrated Child Development Services (ICDS), GoI, 2020-2021." *Budget Briefs*, 12(4). Delhi, India: Centre for Policy Research.

Alex-Petersen, J., P. Lundborg, and D.-O. Rooth (2017). "Long-term effects of childhood nutrition: Evidence from a school lunch reform." (WP No. 11234). Bonn, Germany: IZA.

Andrabi, T. J., J. Das, and A. I. Khwaja (2013). "Students today, teachers tomorrow: Identifying constraints on the provision of education." *Journal of Public Economics* 100, pp. 1–14.

Andrew, A., O. Attanasio, R. Bernal, L. Cardona Sosa, S. Krutikova, and M. Rubio-Codina (2019). "Pre-school quality and child development." (NBER WP 26191). Cambridge, MA.

ASER (2019). "Annual status of education report (rural) 2018." Delhi, India: ASER.

— (2020). "Annual status of education report (rural) 2019: Early years." Delhi, India: ASER.

Bailey, M. J., B. D. Timpe, and S. Sun (2021). "Prep school for poor kids: The long-run impacts of Head Start on human capital and economic self-sufficiency." *American Economic Review* 111 (12), pp. 3963–4001.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). "Remedying education: Evidence from two randomized experiments in India." *The Quarterly Journal of Economics* 122 (3), pp. 1235–1264.

Barham, T., K. Macours, and J. A. Maluccio (2013). "Boys' cognitive skill formation and physical growth: Long-term experimental evidence on critical ages for early childhood interventions." *American Economic Review* 103 (3), pp. 467–71.

Bau, N. and J. Das (2020). "Teacher value-added in a low-income country." *American Economic Journal: Economic Policy* 12 (1), pp. 62–96.

Berlinski, S., S. Galiani, and P. J. Gertler (2009). "The effect of pre-primary education on primary school performance." *Journal of Public Economics* 93 (1), pp. 219–234.

[38] For instance, recent experimental evidence in a different Indian state (Odisha) demonstrates that home-visiting programs designed to support psycho-social stimulation can generate large improvements in child development (Grantham-McGregor et al., 2020). However, these interventions were implemented by an additional staff member employed by a nonprofit agency, and a key barrier to scaling up this approach through the ICDS has been the lack of *anganwadi* staff time to do so.

[39] Studying longer-term persistence of effects would help to account for any adjustments in behavior over time as workers get used to the presence of new AWC staff.

Berlinski, S., S. Galiani, and M. Manacorda (2008). "Giving children a better start: Preschool attendance and school-age profiles." *Journal of Public Economics* 92 (5-6), pp. 1416–1440.

Berlinski, S. and N. Schady (2015). "The early years: Child well-being and the role of public policy." Washington, DC: Inter-American Development Bank.

Blatchford, P. and P. Mortimore (1994). "The issue of class size for young children in schools: what can we learn from research?" *Oxford review of education* 20 (4), pp. 411–428.

Blimpo, M. P., P. Carneiro, P. Jervis, and T. Pugatch (2022). "Improving access and quality in early childhood development programs: Experimental evidence from The Gambia." *Economic Development and Cultural Change* 70 (4), pp. 1479–1529.

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau (2018). "Preschool and parental response in a second best world evidence from a school construction experiment." *Journal of Human Resources* 53 (2), pp. 474–512.

Britto, P. R., S. J. Lye, K. Proulx, A. K. Yousafzai, S. G. Matthews, T. Vaivada, R. Perez-Escamilla, N. Rao, P. Ip, L. C. H. Fernald, et al. (2017). "Nurturing care: promoting early childhood development." *The Lancet* 389 (10064), pp. 91–102.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly Journal of Economics* 126 (4), pp. 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review* 104 (9), pp. 2633–79.

Cole, S., A. Paulson, and G. K. Shastry (2016). "High school curriculum and financial outcomes: The impact of mandated personal finance and mathematics courses." *Journal of Human Resources* 51 (3), pp. 656–698.

Dasgupta, A. and D. Kapur (2020). "The political economy of bureaucratic overload: Evidence from rural development officials in India." *American Political Science Review* 114 (4), pp. 1316–1334.

de Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018). "Double for nothing? Experimental evidence on the impact of an unconditional teacher salary increase on student performance in Indonesia." *The Quarterly Journal of Economics* 133 (2), pp. 993–1039.

Dean, J. and S. Jayachandran (2019). "The impact of early childhood education on child development in rural India." *Unpublished manuscript.* Karnataka, India: J-PAL.

Duflo, E., P. Dupas, and M. Kremer (2011). "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya." *American Economic Review* 101, pp. 1739–1774.

— (2015). "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools." *Journal of Public Economics* 123, pp. 92–110.

Engle, P. L, M. M. Black, J. R. Behrman, M. C. De Mello, P. J. Gertler, L. Kapiriri, R. Martorell, M. E. Young, and International Child Development Steering Group (2007). "Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world." *The Lancet* 369 (9557), pp. 229–242.

Finan, F., B. A. Olken, and R. Pande (2017). "The personnel economics of the developing state." In Banerjee, A. V. and Duflo, E. *Handbook of field experiments (Vol. II).* Oxford, UK: North Holland.

Ganimian, A. J., K. Muralidharan, and C. R. Walters (2020). “Improving school preparedness and child health outcomes through Integrated Child Development Services in Tamil Nadu.” Report to the Government of Tamil Nadu. Delhi, India: J-PAL.

Garces, E., D. Thomas, and J. Currie (2002). “Longer-term effects of Head Start.” *American Economic Review* 92 (4), pp. 999–1012.

Gelli, A., E. Aurino, G. Folson, D. Arhinful, C. Adamba, I. Osei-Akoto, E. Masset, K. Watkins, M. Fernandes, L. Drake, and H. Alderman (2019). “A school meals program implemented at scale in Ghana increases height-for-age during midchildhood in girls and in children from poor households: A cluster randomized trial.” *The Journal of Nutrition* 149 (8), pp. 1434–1442.

GoI (2020). “National Education Policy 2020.” New Delhi, India: Ministry of Human Resource Development, Government of India.

GoTN (2021). “Criteria for the appointment of *anganwadi* worker/mini *anganwadi* worker/helper.” URL: `https://bit.ly/3TDDf4p`. Last accessed: January 20, 2021.

Grantham-McGregor, S. M., A. Adya, O. P. Attanasio, B. Augsburg, J. R. Behrman, B. Caeyers, M. Day, P. Jervis, R. Kochar, P. Makkar, et al. (2020). “Group sessions or home visits for early childhood development in India: a cluster RCT.” *Pediatrics* 146 (6).

Haines, A., D. Sanders, U. Lehmann, A. K. Rowe, J. E. Lawn, S. Jan, D. G. Walker, and Z. A. Bhutta (2007). “Achieving child survival goals: Potential contribution of community health workers.” *The Lancet* 369, pp. 2121–2131.

Hazarika, G. and V. Viren (2013). “The effect of early childhood developmental program attendance on future school enrollment in rural North India.” *Economics of Education Review* 34, pp. 146–161.

Heaver, R. (2002). “India’s Tamil Nadu Nutrition Program: Lessons and Issues in Management and Capacity Development.” Washington, DC: World Bank.

Heckman, J. J. (2012). “Invest in early childhood development: Reduce deficits, strengthen the economy.” *The Heckman Equation* 7, pp. 1–2.

Heckman, J. J., J. L. Garcia, F. Bennhoff, and D. E. Leaf (2021). “The dynastic benefits of early childhood education.” (WP No. 2021-77). Chicago, IL: University of Chicago.

Heckman, J. J. and S. Mosso (2014). “The economics of human development and social mobility.” *Annual Review of Economics* 6, pp. 689–733.

Hendren, N. and B. Sprung-Keyser (2020). “A unified welfare analysis of government policies.” *The Quarterly Journal of Economics* 135 (3), pp. 1209–1318.

Hoddinott, J., H. Alderman, J. R. Behrman, L. Haddad, and S. Horton (2013). “The economic rationale for investing in stunting reduction.” *Maternal & child nutrition* 9, pp. 69–82.

Hoddinott, J., J. Maluccio, J. R. Behrman, R. Martorell, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, and K. M. Yount (2011). “The consequences of early childhood growth failure over the life course.” (WP No. 01073). Washington, DC: IFPRI.

ICDS (2017). “G.O. (MS) No. 110 - Eligibility criteria for the post of AWW, AWH and Mini AWW.” Tamil Nadu Integrated Child Development Services. Department of Social Welfare and Nutritious Meal Programme. Government of Tamil Nadu. Chennai.

Jensen, R. (2012). “Do labor market opportunities affect young women’s work and family decisions? Experimental evidence from India.” *The Quarterly Journal of Economics* 127 (2), pp. 753–792.

Kleiner, M. M. (2000). “Occupational licensing.” *Journal of Economic Perspectives* 14 (4), pp. 189–202.

Kline, P. and C. R. Walters (2016). "Evaluating public programs with close substitutes: The case of Head Start." *The Quarterly Journal of Economics* 131 (4), pp. 1795–1848.
Kling, Jeffrey R and Jeffrey B Liebman (2004). *Experimental analysis of neighborhood effects on youth.* Tech. rep. Working Paper.
Lazear, E. P. (2001). "Educational production." *The Quarterly Journal of Economics* 116 (3), pp. 777–803.
List, J. A. (2021). "The voltage effect: How to make good ideas great and great ideas scale." Penguin.
List, J. A., D. Suskind, and L. H. Supplee (2021). "The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it." Routledge.
Lu, C., M. M. Black, and L. M. Richter (2016). "Risk of poor development in young children in low-income and middle-income countries: An estimation and analysis at the global, regional, and country level." *The Lancet* 4, e916–e922.
Martinez, S., S. Naudeau, and V. Pereira (2017). "Preschool and child development under extreme poverty: Evidence from a randomized experiment in rural Mozambique." (WP No. 8290). Washington, DC: The World Bank.
MHFW (2020). "National family health survey-5 (2019-20). Key indicators from 22 states/UTs from phase I." Delhi, India: Government of India.
Moore, C., E. K. Fletcher, and R. Pande (2018). "Women and work in India: Descriptive evidence and a review of potential policies." *India Policy Forum 2018* 15 (1), pp. 149–216.
Muralidharan, K. (2016). "A new approach to public sector hiring in India for improved service delivery." *India Policy Forum 2015-16*. New Delhi, India: NCAER.
Muralidharan, K. and P. Niehaus (2017). "Experimentation at scale." *Journal of Economic Perspectives* 31 (4), pp. 103–124.
Muralidharan, K., P. Niehaus, and S. Sukhtankar (2021). "General equilibrium effects of (improving) public employment programs: Experimental evidence from India." NBER WP 23838. Cambridge, MA.
Muralidharan, K., M. Romero, and K. Wüthrich (2021). "Factorial designs, model selection, and (incorrect) inference in randomized experiments." NBER WP 26562. Cambridge, MA.
Muralidharan, K., A. Singh, and A. J. Ganimian (2019). "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review* 109 (4), pp. 1426–60.
Muralidharan, K. and V. Sundararaman (2013). "Contract teachers: Experimental evidence from India." (WP No. 19440). Cambridge, MA: NBER.
MWCD (2019). "Anganwadi sevikas." Delhi, India: Ministry of Women and Child Development. URL: `https://bit.ly/3U28fLx`. Last accessed: July 16, 2021.
Nandi, A., A. Ashok, S. Kinra, J. R. Behrman, and R. Laxminarayan (2016). "Early childhood nutrition is positively associated with adolescent educational outcomes: Evidence from the Andhra Pradesh Child and Parents Study." *The Journal of Nutrition* 146 (4), pp. 806–813.
Nandi, A., J. R. Behrman, S. Kinra, and R. Laxminarayan (2018). "Early-life nutrition is associated positively with schooling and labor market outcomes and negatively with marriage rates at age 20–25 years: evidence from the Andhra Pradesh Children and Parents Study in India." *The Journal of Nutrition* 148 (1), pp. 140–146.
Nandi, A., J. R. Behrman, and R. Laxminarayan (2020). "The impact of a national early childhood development program on future schooling attainment: Evidence from Integrated

Child Development Services in India." *Economic Development and Cultural Change* 69 (1), pp. 291–316.

NHSA (2020). "2020 national Head Start profile." Alexandria, VA: National Head Start Association.

PEO (2011). "Evaluation report on Integrated Child Development Services." Delhi, India: Planning Commission, Government of India.

Prasad, V. and D. Sinha (2015). "The reluctant state: Lacunae in current child health and nutrition policies and programmes in India." In *India social development report: Challenges of public health*, pp. 69-81. Oxford, UK: Oxford University Press.

Pritchett, L. (2009). "Is India a flailing state? Detours on the four lane highway to modernization." (WP No. RWP09-013). Cambridge, MA: Harvard Kennedy School.

Ravindran, S. (2020). "Parental investments and early childhood development: Short and long run evidence from India." *Unpublished manuscript*. New York, NY: New York University.

Sen, A. (2001). "Development as freedom." Oxford Paperbacks.

Shonkoff, J. P., D. A. Phillips, et al. (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academies Press.

Singh, A. (2014). "Giving up too early on malnourished children? Catch-up growth and Midday Meals." *Ideas for India*. URL: `https://bit.ly/3FsqCF4`. Last accessed: April 26, 2021.

Singh, A., A. Park, and S. Dercon (2014). "School meals as a safety net: An evaluation of the Midday Meal scheme in India." *Economic Development and Cultural Change* 62 (2), pp. 275–306.

Singh, A., M. Romero, and K. Muralidharan (2022). "Covid-19 learning loss and recovery: Panel data evidence from India." NBER WP 30552. Cambridge, MA.

Sinha, D. (Aug. 2006). "Rethinking ICDS: A rights based perspective." *EPW*.

Sinha, D., M. Gupta, and D. Shriyan (2021). "High risk without recognition: Challenges faced by female front-line workers." *EPW*, May 3, 2021.

Stallings, J. A. and G. G. Mohlman (1990). "Issues in qualitative education research: Observation techniques." In H. J. Walberg & G. D. Haertel (Eds.) *The international encyclopedia of educational evaluation* (pp. 639-644). New York, NY: Pergamon Press.

The Times of India (2021). "Anganwadi workers demand pay hike, permanent jobs." *The Times of India*. February 23, 2021.

Urquiola, Miguel (2006). "Identifying class size effects in developing countries: Evidence from rural Bolivia." *Review of Economics and statistics* 88 (1), pp. 171–177.

Vogl, T. (2012). "Education and health in developing economies." *Encyclopedia of Health Economics* 1453, pp. 246–249.

Working Group for Children Under Six (Dec. 2012). "Strategies for children under six: Updates and recommendations for the 12th plan." Working Group for Children Under Six of the Right to Food Campaign and Jan Swasthya Abhiyan.

World Bank (2018a). "Evaluating integration in the ICDS: Impact evaluation of an AWC-cum-creche pilot in Madhya Pradesh." (Report No: AUS0000253). Washington, DC: World Bank.

— (2018b). "World Development Report 2018: Learning to realize education's promise." Washington, DC: World Bank.

— (2021). "Cost-effective approaches to improve global learning: What does recent evidence tell us are "smart buys" for improving learning in low- and middle-income countries?" Washington, DC: World Bank; FCDO; and Building Evidence in Education.

Table 1: Summary statistics at baseline and randomization balance

	(1) Control	(2) Treatment	(3) Difference	(4) N
A. Anganwadi centers				
No. of children registered at AWC	20.050	19.394	-0.656	320
	[7.062]	[5.669]	(0.705)	
No. of children observed at the AWC	15.006	14.212	-0.794	320
	[6.428]	[5.384]	(0.658)	
Located next to primary school	0.294	0.312	0.019	320
	[0.457]	[0.465]	(0.051)	
Functions jointly with another AWC	0.044	0.087	0.044	320
	[0.205]	[0.283]	(0.027)	
Has electricity connection	0.769	0.762	-0.006	320
	[0.423]	[0.427]	(0.046)	
Has kitchen	0.856	0.875	0.019	320
	[0.352]	[0.332]	(0.037)	
Has toilet	0.688	0.762	0.075	320
	[0.465]	[0.427]	(0.050)	
B. Anganwadi workers				
Age	50.566	48.481	-2.178**	315
	[8.151]	[9.494]	(0.963)	
Passed grade 10 or higher	0.831	0.894	0.062	320
	[0.376]	[0.309]	(0.039)	
Years of experience as AWW	23.094	20.475	-2.619**	320
	[9.811]	[10.696]	(1.118)	
No. days of training last year	6.019	5.308	-0.741	315
	[6.856]	[4.720]	(0.643)	
Received ECE training	0.606	0.663	0.056	320
	[0.490]	[0.474]	(0.053)	
Salary (INR)	8,169.394	7,861.569	-307.825**	320
	[1,162.451]	[1,239.866]	(128.820)	
C. Children				
Female	0.504	0.518	0.013	4,675
	[0.500]	[0.500]	(0.013)	
Age	3.559	3.468	-0.082**	4,661
	[0.848]	[0.868]	(0.033)	
Weight-for-age (WAZ) score	-1.658	-1.599	0.046	4,568
	[1.025]	[1.017]	(0.039)	
Height-for-age (HAZ) score	-1.565	-1.483	0.071	4,528
	[1.470]	[1.358]	(0.051)	
Underweight	0.369	0.350	-0.016	4,568
	[0.483]	[0.477]	(0.017)	
Stunted	0.352	0.334	-0.015	4,528
	[0.478]	[0.472]	(0.016)	
Math (std. proportion-correct score)	-0.018	0.026	0.015	4,675
	[0.992]	[1.004]	(0.046)	
Language (std. proportion-correct score)	-0.026	-0.005	-0.012	4,675
	[0.983]	[0.981]	(0.045)	
Exec. function (std. proportion-correct score)	0.004	0.013	-0.007	4,675
	[0.997]	[1.029]	(0.046)	
Composite score	-0.022	0.023	-0.001	4,675
	[1.433]	[1.448]	(0.068)	

Notes: This table compares the *anganwadi* centers (AWCs), *anganwadi* workers (AWWs), and children in the control and treatment groups at baseline. It shows the means and standard deviations for each group (columns 1-2) and tests for differences between groups including randomization-strata fixed effects (column 3). Panel A displays figures for AWCs, Panel B for AWWs, and Panel C for children. The sample includes all AWCs, AWWs, and children observed at baseline. The composite standardized score is the first principal component from a principal component analysis of scores on all three subjects. Baseline scores are standardized to have mean zero and standard deviation one in the full sample. Standard deviations appear in brackets, and standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2: Follow-up rate in endline assessments

	(1) AWC assessment	(2) HH assessment
A. Treatment		
Treatment	-0.008	-0.022
	(0.016)	(0.018)
N (children)	4675	2336
Control mean	0.328	0.892
B. Treatment and baseline		
Treatment	-0.044	-0.031
	(0.075)	(0.072)
Female	0.017	-0.001
	(0.019)	(0.018)
Age (at baseline)	-0.203***	-0.016
	(0.013)	(0.014)
WAZ score (at baseline)	-0.055***	0.034**
	(0.013)	(0.016)
HAZ score (at baseline)	-0.013	-0.050***
	(0.009)	(0.016)
Composite score (at baseline)	-0.015**	0.015
	(0.007)	(0.010)
Female × Treatment	-0.037	-0.014
	(0.026)	(0.027)
Age × Treatment	0.009	-0.001
	(0.018)	(0.020)
WAZ × Treatment	0.001	-0.020
	(0.020)	(0.022)
HAZ × Treatment	-0.008	0.015
	(0.015)	(0.021)
Composite score × Treatment	0.003	0.010
	(0.011)	(0.013)
N (children)	4521	2271
F-ratio (Interactions)	0.602	0.389
P-value	0.698	0.857

Notes: The table shows estimates from regressions predicting follow-up status in the endline assessments conducted in AWCs (column 1) and households (column 2). Follow-up is defined as having an observed test score at endline. The sample includes children present at AWCs at baseline. Panel A regresses follow-up on treatment status, and Panel B regresses follow-up on treatment status interacted with baseline characteristics. Both panels include randomization-strata fixed effects. Column 1 includes all children present at baseline, while column 2 includes children sampled for the household survey and weights by the inverse of the survey sampling weights. Standard errors (clustered by AWC) appear in parentheses. The F- and p-values refer to a test of joint significance for all interaction terms. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Impact on attendance and punctuality from unannounced observations

	(1)	(2)	(3)
	AWCs		Impact on AWCs
A. Center-level impacts	Control	Treatment	Col. (2)-(1)
Share of centers that were...			
...open by AWC opening time (9am)	0.400	0.488	0.049
	[0.493]	[0.503]	(0.081)
...open by PSE start time (10am)	0.875	0.962	0.089**
	[0.333]	[0.191]	(0.045)
...closed	0.125	0.038	-0.089**
	[0.333]	[0.191]	(0.045)
N (centers)	80	80	160

	(1)	(2)	(3)	(4)
	AWWs		Impact on AWWs	ECE facilitators
B. Worker-level impacts	Control	Treatment	Col. (2)-(1)	Treatment
Share of workers who...				
...arrived by AWC opening time (9am)	0.125	0.237	0.100	0.213
	[0.333]	[0.428]	(0.064)	[0.412]
...arrived by PSE start time (10am)	0.800	0.900	0.107*	0.913
	[0.403]	[0.302]	(0.059)	[0.284]
...were absent	0.200	0.100	-0.107*	0.087
	[0.403]	[0.302]	(0.059)	[0.284]
N (centers)	80	80	160	80

Notes: This table compares average attendance and punctuality of AWWs in control and treatment AWCs and of AWWs and facilitators in treatment AWCs, based on unannounced visits about a year after the rollout of the intervention (February 2018). Panel A displays results for AWCs and Panel B shows results for frontline workers (AWWs or facilitators). Impact estimates come from regressions of each variable on a treatment indicator with controls for AWW education and experience and indicators for missing values. Standard deviations appear in brackets, and standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Impact on overall time allocation from unannounced observations

	(1) ECE facilitators	(2) AWWs	(3)	(4) Impact on AWWs	(5) AWWs & facilitators	(6) Impact on AWCs
Minutes per day...	Treatment	Control	Treatment	Col. (3)-(2)	Col. (1)+(3)	Col. (5)-(2)
...on pre-school education	57.450	38.400	18.150	-19.014***	75.600	38.908***
	[31.530]	[29.665]	[21.432]	(4.183)	[37.092]	(5.386)
...on administrative tasks	19.650	21.900	35.100	13.091***	54.750	32.780***
	[17.834]	[22.084]	[26.715]	(4.022)	[34.519]	(4.712)
...on health and nutrition tasks	5.550	5.700	10.800	5.691***	16.350	11.317***
	[8.527]	[9.917]	[14.616]	(1.977)	[18.338]	(2.315)
...off duty	26.850	30.000	43.950	13.131***	70.800	39.751***
	[24.019]	[26.799]	[30.366]	(4.743)	[45.808]	(6.045)
...absent	10.500	24.000	12.000	-12.899*	22.500	-2.755
	[34.122]	[48.303]	[36.227]	(7.094)	[60.661]	(8.935)
N (centers)	80	80	80	160	80	160

Notes: This table compares the average time allocation of AWWs in control and treatment AWCs and of AWWs and facilitators in treatment AWCs, measured in unannounced visits about a year after the rollout of the intervention (February 2018). Time allocation was recorded from 10am to 12pm, during the time officially designated for preschool education. Impact estimates come from regressions of each variable on a treatment indicator with controls for AWW education and experience and indicators for missing values. Standard deviations appear in brackets, and standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Impact on endline assessments (standardized scores)

	(1) Math	(2) Language	(3) Executive function	(4) Composite score
A. Complete sample				
AWC assessments (N=1514)				
Treatment	0.291***	0.459***	0.180***	0.288***
	(0.061)	(0.081)	(0.052)	(0.058)
HH assessments (N=2075)				
Treatment	0.130***	0.101**	0.057	0.110**
	(0.049)	(0.051)	(0.042)	(0.045)
B. Common sample				
AWC assessments (N=791)				
Treatment	0.311***	0.460***	0.205***	0.315***
	(0.075)	(0.095)	(0.068)	(0.071)
HH assessments (N=791)				
Treatment	0.290***	0.361***	0.158**	0.291***
	(0.080)	(0.091)	(0.062)	(0.071)
P-value (AWC = HH)	0.935	0.330	0.407	0.726

Notes: The table shows the impact of the intervention on assessments of math, language, and executive function after two years. Estimates come from regressions of endline test scores on a treatment indicator with controls for randomization strata and baseline characteristics. Endline scores are standardized to have mean zero and standard deviation one in the control group. The composite score is the first principal component of math, language, and executive function scores. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline assessments. Panel B displays results for all children who participated in the baseline and both endline assessments. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. The last row displays the p-value testing the null hypothesis that the treatment effects across both assessments in Panel B are equal. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: Impact on endline WAZ and HAZ scores (complete sample)

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
AWC measurements (N=1538)			
Treatment	0.096***	-0.018	-0.031**
	(0.033)	(0.017)	(0.012)
Control mean	-1.762	0.384	0.091
HH measurements (N=2016)			
Treatment	0.049	-0.015	-0.006
	(0.032)	(0.018)	(0.011)
Control mean	-1.553	0.321	0.075
	(1) HAZ score	(2) Stunted (HAZ<-2)	(3) Severely stunted (HAZ<-3)
AWC measurements (N=1389)			
Treatment	0.092**	-0.048**	-0.023**
	(0.044)	(0.022)	(0.011)
Control mean	-1.492	0.291	0.057
HH measurements (N=1990)			
Treatment	0.014	-0.027	-0.010
	(0.054)	(0.017)	(0.007)
Control mean	-1.167	0.205	0.040

Notes: The table shows the impact of the intervention on children's weight-for-age and height-for-age z-scores (WAZ and HAZ, respectively), the share of underweight (WAZ below -2) and stunted (HAZ below -2) children, and the share of severely underweight (WAZ below -3) and severely stunted (HAZ below -3) children. Estimates come from regressions of WAZ/HAZ outcomes on a treatment indicator with controls for randomization strata and baseline characteristics. Both panels display results for all children with baseline measurements and endline measurements in either the AWC or household (HH) survey. Estimates for the HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC sample do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: Cost/benefit analysis

Parameter	Source	HH sample
A. Projecting future earnings		
Labor force participation rate	LFP for rural population of TN aged 15+, 2011-2012 NSS[a]	0.52
Current average daily wage	Average wage for rural workers aged 15+, 2011-2012 NSS[b]	268
Days worked per year when in labor force	Assumption	225
Current annual earnings when in labor force	Calculation	60,300
Annual real wage growth	Assumption[c]	0.05
Working age	Assumption	22-65
Discount rate	Assumption	0.03
Average PDV of lifetime earnings at age 4	Calculation	3,622,089
B. Experimental impacts		
Test-score effect (std.)	Experimental estimate[d]	0.11
Earnings gain per std. dev. of test scores	Literature estimates linking test scores to earnings[e]	0.13
Predicted PDV earnings gain per child	Calculation	51,786
C. Benefit-cost ratio		
Children per center per year	Experimental data	14
Cohort size adjustment factor	Assumption[f]	1.33
Predicted benefit per center	Calculation	964,439
Program cost per center	Government order	74,478
Benefit-cost ratio	Calculation	12.9

Notes: This table reports a cost benefit analysis of the ECCE facilitator intervention based on projected impacts on adult earnings. Panel A lists the parameters necessary to project the present discounted value (PDV) of lifetime earnings for children in Tamil Nadu. Panel B lists parameters and assumptions necessary to predict the increase in earnings generated by the ECCE facilitator intervention for each child based on the programs' test score effects. Panel C combines this projection with program costs to produce a benefit/cost ratio. Column (1) uses the full household sample, while column (2) assumes all benefits accrue to children in the AWC sample.

a NSS Report No. 554 (July 2011-June 2012), statement 4.1.2.

b NSS Report No. 554 (July 2011-June 2012), statement 5.13.1.

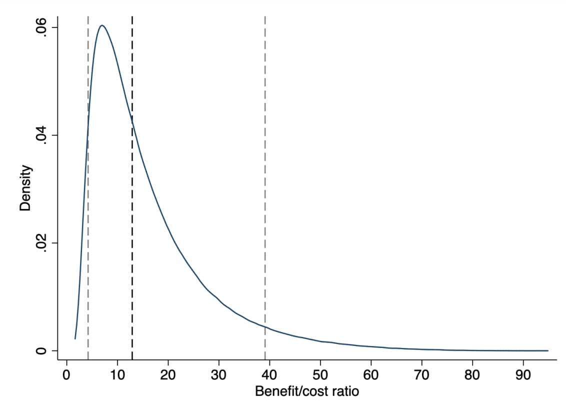
c MOSPI states that real gross state domestic product in Tamil Nadu grew 6.4% per year from 2011 to 2017.

d Table 5, Panel A, column 4.

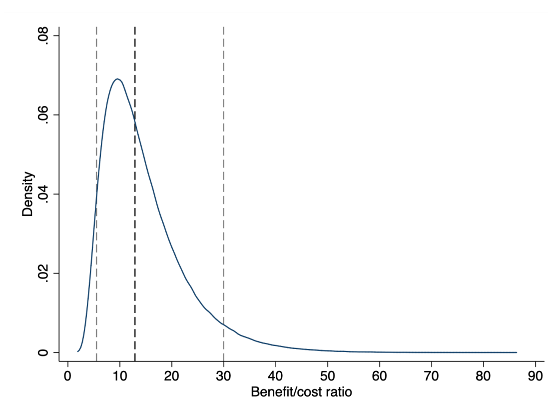
e Preferred estimate from Kline and Walters (2016), Appendix Table A.IV.

f Assuming 25% turnover per year, the experimental data understate the number of children served over two years by 33%.

Figure 1: Benefit-cost sensitivity analysis



Uniform parameters (mean = 15.8, median = 12.5)



Truncated normal parameters (mean = 14.6, median = 12.7)

Notes: This figure explores the sensitivity of the ECCE facilitator benefit/cost ratio to parameter values. Estimates are based on test-score gains for the household sample as in column (1) of Table 5. We obtain a distribution of benefit/cost ratios by drawing each parameter calibrated from other data sources from a range of possible values, with the preferred values from Table 5 in the middle of each range. Days worked in the labor market range from 200 to 250. The wage growth rate ranges from 3 percent to 7 percent. The proportionate increase in earnings per standard deviation of test scores ranges from 7 percent to 19 percent. The discount rate ranges from 1.5 to 4.5 percent. The left-hand plot draws each of these parameters from an independent uniform distribution, while the right-hand plot draws each parameter from an independent truncated normal distribution with mean in the middle of the range and standard deviation 1/4 of the width of the range. Results come from fitting kernel densities to 1,000,000 draws of the parameters, excluding values over 90 for visual presentation. Gray lines indicate 5th and 95th percentiles, and black vertical lines denote our preferred estimates from Table 7.

Appendix A Additional figures and tables

Table A.1: Summary statistics on facilitators from intervention monitoring

	(1)
Had ECE facilitator at AWC	0.981
	[0.136]
Age	28.816
	[3.842]
Received training	0.956
	[0.205]
No. days of training	6.190
	[1.281]
Days since being hired	134.548
	[80.935]
Has ECE activities register	0.788
	[0.410]
ECE activities register is updated	0.712
	[0.454]
N (centers)	160

Notes: This table shows descriptive statistics for facilitators in treatment centers for all variables collected in the first round of intervention monitoring, from April to May of 2017. All intervention-monitoring visits were pre-scheduled. Standard deviations appear in brackets.

Table A.2: Impact on time allocation to health and nutrition from unannounced observations

	(1)	(2)	(3)	(4)	(5)	(6)
	ECE facilitators	AWWs		Impact on AWWs	AWWs & facilitators	Impact on AWCs
Minutes per day...	Treatment	Control	Treatment	Col. (3)-(2)	Col. (1)+(3)	Col. (5)-(2)
...preparing or serving food	2.850	1.800	5.850	4.622***	8.700	7.552***
	[6.681]	[5.088]	[12.664]	(1.506)	[15.509]	(1.786)
...assisting children to use toilet	2.400	2.550	1.650	-0.846	4.050	1.562*
	[4.830]	[4.940]	[4.159]	(0.756)	[5.710]	(0.873)
...on health-related activities	0.300	1.350	3.300	1.915*	3.600	2.202**
	[1.885]	[6.331]	[6.607]	(1.035)	[7.245]	(1.086)
N (centers)	80	80	80	160	80	160

Notes: This table compares the time allocation of AWWs in control and treatment AWCs and of AWWs and facilitators in treatment AWCs, measured in unannounced visits about a year after the rollout of the intervention (February 2018). Time allocation was recorded from 10am to 12pm, during the time officially designated for preschool education. Standard deviations appear in brackets and standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.3: Impact on endline assessments with inverse-probability weights (standardized scores)

	(1) Math	(2) Language	(3) Executive function	(4) Composite score
A. Complete sample				
AWC assessments (N=1514)				
Treatment	0.280***	0.386***	0.168***	0.266***
	(0.073)	(0.084)	(0.062)	(0.067)
HH assessments (N=2075)				
Treatment	0.109**	0.081	0.049	0.086*
	(0.049)	(0.052)	(0.042)	(0.045)
B. Common sample				
AWC assessments (N=791)				
Treatment	0.323***	0.392***	0.223***	0.323***
	(0.090)	(0.109)	(0.074)	(0.081)
HH assessments (N=791)				
Treatment	0.283***	0.367***	0.130*	0.275***
	(0.092)	(0.097)	(0.068)	(0.080)

Notes: The table shows the impact of the intervention on assessments of math, language, and executive function after two years, weighted by the inverse of the predicted probability of participating in the endline assessments. This probability is predicted using a probit model with baseline scores in language, math, and executive function, sex of the child, and indicator variables for centers as predictors. Estimates come from regressions of endline test scores on a treatment indicator with controls for randomization strata and baseline characteristics. Endline scores are standardized to have mean zero and standard deviation one in the control group. The composite score is the first principal component of math, language, and executive function scores. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline assessments. Panel B displays results for all children who participated in the baseline and both endline assessments. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. The last row displays the p-value testing the null hypothesis that the treatment effects across both assessments in Panel B are equal. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.4: Heterogeneous impacts on endline assessments (standardized scores)

	(1) Baseline WAZ score	(2) Female	(3) Age (months)	(4) Mother education	(5) AWW vacancy	(6) Class size	(7) Baseline composite score
A. Complete sample							
AWC measurements							
Treatment	0.249**	0.201***	0.127	0.343***	0.271***	0.401	0.273***
	(0.124)	(0.072)	(0.264)	(0.092)	(0.064)	(0.432)	(0.070)
Covariate	0.012	0.010	0.043***	0.110	0.050	0.049	
	(0.039)	(0.065)	(0.005)	(0.071)	(0.102)	(0.098)	
Interaction	-0.019	0.172*	0.005	-0.069	0.054	-0.040	-0.010
	(0.056)	(0.097)	(0.007)	(0.104)	(0.146)	(0.151)	(0.067)
N (children)	1485	1514	1511	1514	1514	1514	1514
HH measurements							
Treatment	0.050	-0.073	0.148	0.097	0.105*	0.256	0.107**
	(0.094)	(0.058)	(0.163)	(0.059)	(0.054)	(0.302)	(0.045)
Covariate	0.069**	-0.073	0.042***	0.010	0.114	0.009	
	(0.033)	(0.059)	(0.003)	(0.057)	(0.095)	(0.088)	
Interaction	-0.036	0.356***	-0.001	0.027	-0.010	-0.053	0.014
	(0.045)	(0.084)	(0.004)	(0.079)	(0.112)	(0.110)	(0.033)
N (children)	2040	2075	2074	2075	2075	2075	2075
B. Common sample							
AWC measurements							
Treatment	0.187	0.237**	-0.055	0.364***	0.280***	0.495	0.315***
	(0.161)	(0.093)	(0.297)	(0.096)	(0.084)	(0.540)	(0.087)
Covariate	0.065	0.094	0.042***	0.125	-0.071	-0.085	
	(0.050)	(0.081)	(0.006)	(0.088)	(0.128)	(0.128)	
Interaction	-0.058	0.140	0.010	-0.087	0.146	-0.068	0.000
	(0.072)	(0.127)	(0.008)	(0.133)	(0.160)	(0.191)	(0.091)
HH measurements							
Treatment	0.282	0.261***	-0.279	0.208**	0.278***	-0.565	0.315***
	(0.180)	(0.099)	(0.357)	(0.095)	(0.078)	(0.642)	(0.087)
Covariate	0.059	0.164*	0.032***	-0.024	-0.051	-0.130	
	(0.057)	(0.093)	(0.007)	(0.094)	(0.174)	(0.140)	
Interaction	0.001	0.046	0.016	0.194	0.061	0.305	0.000
	(0.079)	(0.146)	(0.010)	(0.141)	(0.210)	(0.224)	(0.091)
N (children)	780	791	791	791	791	791	791

Notes: The table shows the impact of the intervention on assessments of math, language, and executive function after two years, by seven variables collected at baseline. Estimates come from regressions of endline test scores on a treatment indicator, an indicator for the baseline variable, and their interaction, with controls for randomization strata, baseline child and center scores, AWW experience and education, and missingness. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline assessments. Panel B displays results for all children who participated in the baseline and both endline assessments. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. Class size is the natural logarithm of the number of children observed at the center at baseline. The composite standardized score is the first principal component from a principal-component analysis of scores on all three subjects. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Impact on endline WAZ and HAZ scores (commmon sample)

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
AWC measurements (N=790)			
Treatment	0.121***	-0.022	-0.045**
	(0.036)	(0.025)	(0.020)
Control mean	-2.002	0.519	0.138
HH measurements (N=790)			
Treatment	0.074*	-0.007	-0.021
	(0.041)	(0.030)	(0.021)
Control mean	-1.998	0.514	0.151
P-value (AWC = HH)	0.159	0.538	0.120
	(1) HAZ score	(2) Stunted (HAZ<-2)	(3) Severely stunted (HAZ<-3)
AWC measurements (N=724)			
Treatment	0.153***	-0.090***	-0.042***
	(0.054)	(0.028)	(0.016)
Control mean	-1.674	0.375	0.080
	(1)	(2)	(3)
HH measurements (N=724)			
Treatment	0.101*	-0.053*	-0.026
	(0.055)	(0.030)	(0.016)
Control mean	-1.624	0.367	0.075
P-value (AWC = HH)	0.183	0.071	0.268

Notes: The table shows the impact of the intervention on children's weight-for-age and height-for-age z-scores (WAZ and HAZ, respectively), the share of underweight (WAZ below -2) and stunted (HAZ below -2) children, and the share of severely underweight (WAZ below -3) and severely stunted (HAZ below -3) children. Estimates come from regressions of WAZ/HAZ outcomes on a treatment indicator with controls for randomization strata and baseline characteristics. Both panels display results for children with baseline and both endline measurements. Estimates do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. The last row of each panel displays the p-value testing the null hypothesis that the treatment effects across both assessments are equal. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.6: Heterogeneous impacts on endline WAZ scores

	(1) Baseline	(2) Female	(3) Age (months)	(4) Mother education	(5) AWW vacancy	(6) Class size	(7) Baseline score
A. Complete sample							
AWC measurements							
Treatment	0.323***	0.085*	0.326	0.059	0.100***	0.190	0.102**
	(0.115)	(0.044)	(0.235)	(0.050)	(0.037)	(0.276)	(0.042)
Covariate	0.693***	-0.078*	0.005	0.022	0.007	0.041	-0.064**
	(0.046)	(0.042)	(0.004)	(0.051)	(0.063)	(0.066)	(0.030)
Interaction	0.128**	0.021	-0.006	0.056	-0.015	-0.033	0.028
	(0.055)	(0.058)	(0.006)	(0.062)	(0.087)	(0.096)	(0.040)
N (children)	1538	1538	1538	1538	1538	1538	1509
HH measurements							
Treatment	0.154	0.049	0.123	0.057	0.056	0.197	0.057*
	(0.094)	(0.044)	(0.136)	(0.042)	(0.035)	(0.248)	(0.032)
Covariate	0.777***	0.070	-0.001	0.029	0.107	-0.067	-0.000
	(0.034)	(0.052)	(0.002)	(0.044)	(0.074)	(0.062)	(0.023)
Interaction	0.067	-0.005	-0.002	-0.014	-0.046	-0.056	0.021
	(0.050)	(0.070)	(0.003)	(0.065)	(0.073)	(0.087)	(0.028)
N (children)	2016	2016	2016	2016	2016	2016	1984
B. Common sample							
AWC measurements							
Treatment	0.409***	0.114**	0.161	0.085*	0.130***	0.522*	0.145***
	(0.120)	(0.052)	(0.227)	(0.048)	(0.042)	(0.299)	(0.043)
Covariate	0.733***	-0.055	-0.001	-0.022	0.108	0.037	-0.047
	(0.042)	(0.055)	(0.005)	(0.064)	(0.078)	(0.070)	(0.035)
Interaction	0.141***	0.015	-0.001	0.079	-0.057	-0.144	0.046
	(0.052)	(0.074)	(0.006)	(0.078)	(0.094)	(0.105)	(0.044)
HH measurements							
Treatment	0.265**	0.045	-0.024	0.045	0.047	0.343	0.098**
	(0.121)	(0.062)	(0.249)	(0.053)	(0.045)	(0.313)	(0.046)
Covariate	0.766***	0.014	-0.003	0.030	0.054	0.137*	-0.037
	(0.048)	(0.062)	(0.005)	(0.055)	(0.074)	(0.081)	(0.036)
Interaction	0.105*	0.053	0.003	0.073	0.087	-0.093	0.049
	(0.060)	(0.083)	(0.006)	(0.078)	(0.094)	(0.113)	(0.043)
N (children)	790	790	790	790	790	790	779

Notes: The table shows the impact of the intervention on children's weight-for-age z-scores (WAZ) after two years, by seven variables collected at baseline. Estimates come from regressions of endline scores on a treatment indicator, an indicator for the baseline variable, and their interaction, with controls for randomization strata, baseline child and center scores, AWW experience and education, and missingness. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline measurements. Panel B displays results for all children who participated in the baseline and both endline measurements. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. Class size is the natural logarithm of the number of children observed at the center at baseline. The composite standardized score is the first principal component from a principal-component analysis of scores on all three subjects. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.7: Impact on endline assessments by AWC attendance (standardized scores)

	(1) Math	(2) Language	(3) Executive function	(4) Composite score
A. Attends AWC				
HH assessments (N=1129)				
Treatment	0.160***	0.224***	0.134**	0.190***
	(0.057)	(0.055)	(0.055)	(0.051)
Control mean	-0.433	-0.417	-0.259	-0.402
B. Does not attend				
HH assessments (N=946)				
Treatment	0.061	-0.066	-0.031	-0.007
	(0.062)	(0.067)	(0.057)	(0.060)
Control mean	0.829	0.783	0.468	0.752

Notes: The table shows the impact of the intervention on assessments of math, language, and executive function after two years, by whether they were found at the AWC at endline. Estimates come from regressions of endline test scores on a treatment indicator with controls for randomization strata and baseline characteristics. Endline scores are standardized to have mean zero and standard deviation one in the control group. The composite score is the first principal component of math, language, and executive function scores. Panel A displays results for all children who attended AWCs at endline. Panel B displays results for all children who did not attend AWCs at endline. Both sets of estimates weight by the inverse sampling probability for the HH survey. All specifications control for a baseline measure of the dependent variable and AWW experience and education. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.8: Impact on endline WAZ and HAZ scores with inverse-probability weights (complete sample)

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
AWC measurements (N=1513)			
Treatment	0.100**	-0.015	-0.031***
	(0.042)	(0.020)	(0.012)
Control mean	-1.762	0.384	0.091
HH measurements (N=1997)			
Treatment	0.074*	-0.015	-0.002
	(0.041)	(0.017)	(0.010)
Control mean	-1.553	0.321	0.075

	(1) HAZ score	(2) Stunted (HAZ<-2)	(3) Severely stunted (HAZ<-3)
AWC measurements (N=1388)			
Treatment	0.093*	-0.051**	-0.021**
	(0.054)	(0.022)	(0.010)
Control mean	-1.492	0.291	0.057
HH measurements (N=1988)			
Treatment	0.032	-0.032**	-0.008
	(0.061)	(0.016)	(0.007)
Control mean	-1.167	0.205	0.040

Notes: The table shows the impact of the intervention on children's weight-for-age and height-for-age z-scores (WAZ and HAZ, respectively), the share of underweight (WAZ below -2) and stunted (HAZ below -2) children, and the share of severely underweight (WAZ below -3) and stunted (HAZ below -3) children, weighted by the inverse of the predicted probability of participating in the endline assessments. This probability is predicted using a probit model with baseline WAZ and HAZ scores, sex of the child, and indicator variables for centers as predictors (mean-inputing missing baseline scores). Estimates come from regressions of WAZ/HAZ outcomes on a treatment indicator with controls for randomization strata and baseline characteristics. Panel A displays results for all children with baseline measurements and endline measurements in either the AWC or household (HH) survey. Panel B displays results for children with baseline and both endline measurements. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. The last row displays the p-value testing the null hypothesis that the treatment effects across both assessments in Panel B are equal. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.9: Heterogeneous impacts on endline HAZ scores

	(1) Baseline	(2) Female	(3) Age (months)	(4) Mother education	(5) AWW vacancy	(6) Class size	(7) Baseline score
A. Complete sample							
AWC measurements							
Treatment	0.316*	0.086	0.286	0.148**	0.074	-0.109	0.112**
	(0.178)	(0.059)	(0.294)	(0.069)	(0.051)	(0.398)	(0.049)
Covariate	0.481***	-0.006	0.009*	0.152**	-0.019	-0.025	-0.066*
	(0.069)	(0.053)	(0.005)	(0.064)	(0.085)	(0.085)	(0.034)
Interaction	0.130	0.011	-0.005	-0.069	0.074	0.072	0.058
	(0.090)	(0.075)	(0.008)	(0.079)	(0.112)	(0.139)	(0.040)
N (children)	1389	1389	1389	1389	1389	1389	1362
HH measurements							
Treatment	-0.205	-0.011	-0.342	0.117*	0.082	0.071	0.023
	(0.211)	(0.078)	(0.327)	(0.063)	(0.051)	(0.361)	(0.053)
Covariate	0.609***	-0.080	-0.002	0.147**	0.377**	-0.015	-0.017
	(0.053)	(0.074)	(0.005)	(0.065)	(0.162)	(0.083)	(0.030)
Interaction	-0.149	0.053	0.009	-0.191**	-0.314**	-0.021	0.089***
	(0.126)	(0.097)	(0.007)	(0.097)	(0.149)	(0.128)	(0.033)
N (children)	1990	1990	1990	1990	1990	1990	1958
B. Common sample							
AWC measurements							
Treatment	0.409**	0.149**	0.666**	0.155**	0.143**	0.234	0.159***
	(0.196)	(0.073)	(0.309)	(0.067)	(0.057)	(0.418)	(0.055)
Covariate	0.553***	0.029	0.011**	0.044	0.058	0.036	0.004
	(0.057)	(0.068)	(0.005)	(0.072)	(0.121)	(0.099)	(0.036)
Interaction	0.133	0.005	-0.014*	0.004	0.025	-0.028	0.001
	(0.089)	(0.096)	(0.008)	(0.097)	(0.134)	(0.147)	(0.044)
HH measurements							
Treatment	0.347*	0.098	0.941*	0.154**	0.086	1.019**	0.123**
	(0.197)	(0.074)	(0.483)	(0.077)	(0.057)	(0.438)	(0.058)
Covariate	0.523***	0.021	0.024**	0.199**	0.010	0.187*	-0.002
	(0.066)	(0.090)	(0.012)	(0.092)	(0.124)	(0.109)	(0.053)
Interaction	0.132	0.005	-0.022*	-0.082	0.053	-0.325**	0.030
	(0.104)	(0.115)	(0.013)	(0.130)	(0.145)	(0.154)	(0.061)
N (children)	724	724	724	724	724	724	713

Notes: The table shows the impact of the intervention on children's height-for-age z-scores (HAZ) after two years, by seven variables collected at baseline. Estimates come from regressions of endline scores on a treatment indicator, an indicator for the baseline variable, and their interaction, with controls for randomization strata, baseline child and center scores, AWW experience and education, and missingness. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline measurements. Panel B displays results for all children who participated in the baseline and both endline measurements. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. Class size is the natural logarithm of the number of children observed at the center at baseline. The composite standardized score is the first principal component from a principal-component analysis of scores on all three subjects. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.10: Regression-based association between nutrition status and learning outcomes

	(1) Math	(2) Language	(3) Executive function	(4) Composite score
A. Weight-for-age (N=4568)				
Weight-for-age (WAZ) score	0.114***	0.108***	0.092***	0.182***
Underweight (WAZ<-2)	-0.137***	-0.156***	-0.140***	-0.250***
Severely underweight (WAZ<-3)	-0.221***	-0.176***	-0.276***	-0.382***
B. Height-for-age (N=4528)				
Height-for-age (HAZ) score	0.096***	0.092***	0.083***	0.157***
Stunted (HAZ<-2)	-0.192***	-0.228***	-0.239***	-0.378***
Severely stunted (HAZ<-3)	-0.288***	-0.254***	-0.354***	-0.510***

Notes: The table shows the association between standardized learning outcomes and WAZ and HAZ indicators at baseline. Estimates come from regression of achievement on nutrition variables, with controls for randomization strata, clustering standard errors by AWC. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.11: Impacts on endline WAZ scores by AWC attendance

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
A. Attends AWC			
HH measurements (N=1081)			
Treatment	0.113***	-0.011	-0.009
	(0.039)	(0.024)	(0.014)
Control mean	-1.699	0.354	0.084
B. Does not attend			
HH measurements (N=909)			
Treatment	-0.015	-0.004	0.000
	(0.049)	(0.025)	(0.014)
Control mean	-1.385	0.283	0.064

Notes: The table shows the impact of the intervention on children's weight-for-age z-scores (WAZ) after two years, by whether they were found at the AWC at endline. Estimates come from regressions of endline measurements on a treatment indicator with controls for randomization strata and baseline characteristics. Panel A displays results for all children who attended AWCs at endline. Panel B displays results for all children who did not attend AWCs at endline. Both sets of estimates weight by the inverse sampling probability for the HH survey. All specifications control for a baseline measure of the dependent variable and AWW experience and education. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.12: Impacts on endline HAZ scores by AWC attendance

	(1) HAZ score	(2) Stunted (HAZ<-2)	(3) Severely stunted (HAZ<-3)
A. Attends AWC			
HH measurements (N=1081)			
Treatment	0.088	-0.041*	-0.015
	(0.056)	(0.024)	(0.012)
Control mean	-1.369	0.240	0.053
B. Does not attend			
HH measurements (N=909)			
Treatment	-0.014	-0.014	-0.004
	(0.083)	(0.022)	(0.009)
Control mean	-0.932	0.165	0.025

Notes: The table shows the impact of the intervention on children's height-for-age z-scores (HAZ) after two years, by whether they were found at the AWC at endline. Estimates come from regressions of endline measurements on a treatment indicator with controls for randomization strata and baseline characteristics. Panel A displays results for all children who attended AWCs at endline. Panel B displays results for all children who did not attend AWCs at endline. Both sets of estimates weight by the inverse sampling probability for the HH survey. All specifications control for a baseline measure of the dependent variable and AWW experience and education. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

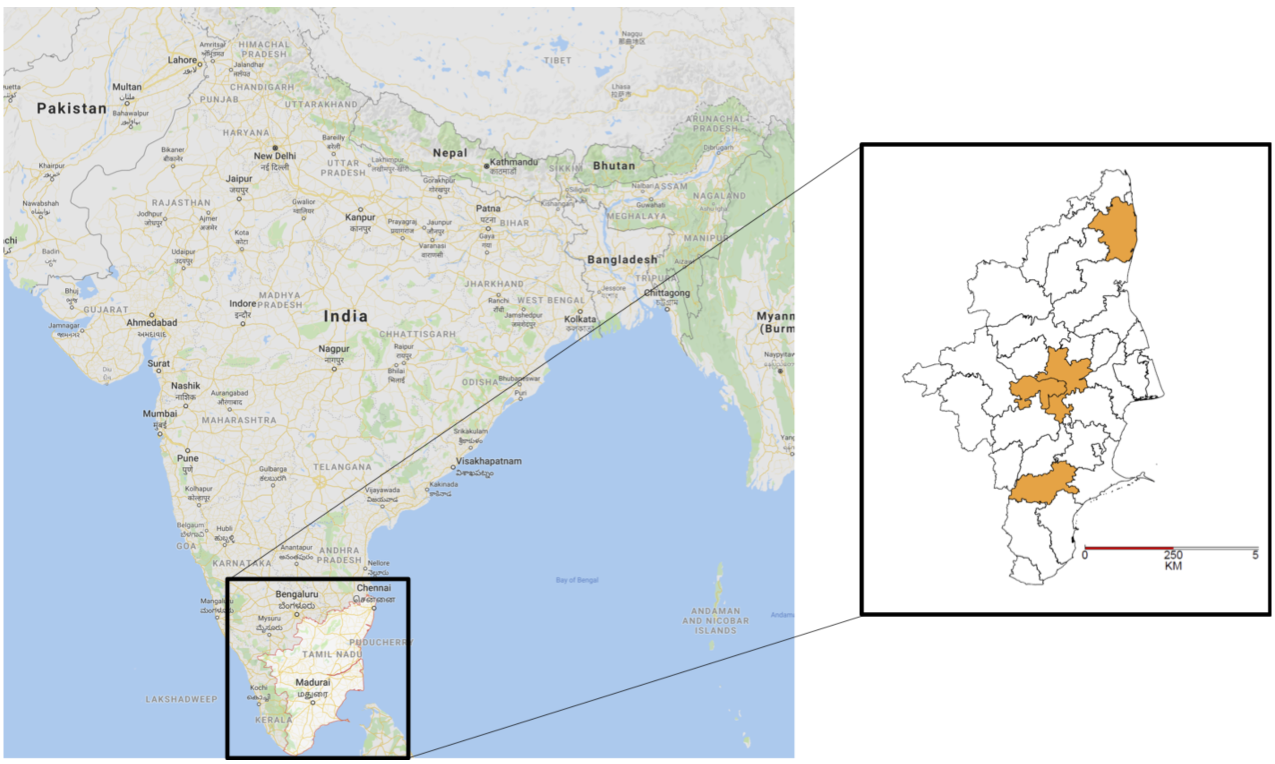
Table A.13: Cost/benefit analysis including nutrition benefits

Parameter	Source	(1) HH sample, test scores	(2) AWC sample, test scores	(3) AWC sample, tests + HAZ	(4) AWC sample, tests + stunting
A. Projecting future earnings					
Labor force participation rate	LFP for rural population of TN aged 15+, 2011-2012 NSS[a]	0.52			
Current average daily wage	Average wage for rural workers aged 15+, 2011-2012 NSS[b]	268			
Days worked per year when in labor force	Assumption	250			
Current annual earnings when in labor force	Calculation	67,000			
Annual real wage growth	Assumption[c]	0.05			
Working age	Assumption	22-65			
Discount rate	Assumption	0.03			
Average PDV of lifetime earnings at age 4	Calculation	3,622,089			
B. Experimental impacts					
Test-score effect (std.)	Experimental estimate[d]	0.11	0.29	0.29	0.29
Earnings gain per std. dev. of test scores	Literature estimates linking test scores to earnings[e]	0.13	0.13	0.13	0.13
Nutrition effect (HAZ or stunting)	Experimental estimate[f]			0.09	-0.05
Earnings gain per unit of nutrition	Literature estimates linking nutrition to earnings[g]			0.20	-0.66
Test-score gain per unit of nutrition	Observational correlation[h]			0.14	-0.37
Predicted PDV earnings gain per child	Calculation	51,796	136,553	195,817	247,488
C. Benefit-cost ratio					
Children per center per year	Experimental data	14	5	5	5
Cohort size adjustment factor	Assumption[i]	1.33	1.33	1.33	1.33
Predicted benefit per center	Calculation	964,439	908,076	1,302,185	1,645,797
Program cost per center	Government order	74,478	74,478	74,478	74,478
Benefit-cost ratio	Calculation	12.9	12.2	17.5	22.1

Notes: This table reports a cost benefit analysis of the ECCE facilitator intervention based on projected impacts on adult earnings. Panel A lists the parameters necessary to project the present discounted value (PDV) of lifetime earnings for children in Tamil Nadu. Panel B lists parameters and assumptions necessary to predict the increase in earnings generated by the ECCE facilitator intervention for each child. Panel C combines this projection with program costs to produce a benefit/cost ratio. Column 1 repeats the results from Table 5, which project benefits using test score gains in the household sample. Column 2 uses test score gains in the AWC sample, and columns 3 and 4 add projected benefits based on nutrition gains.

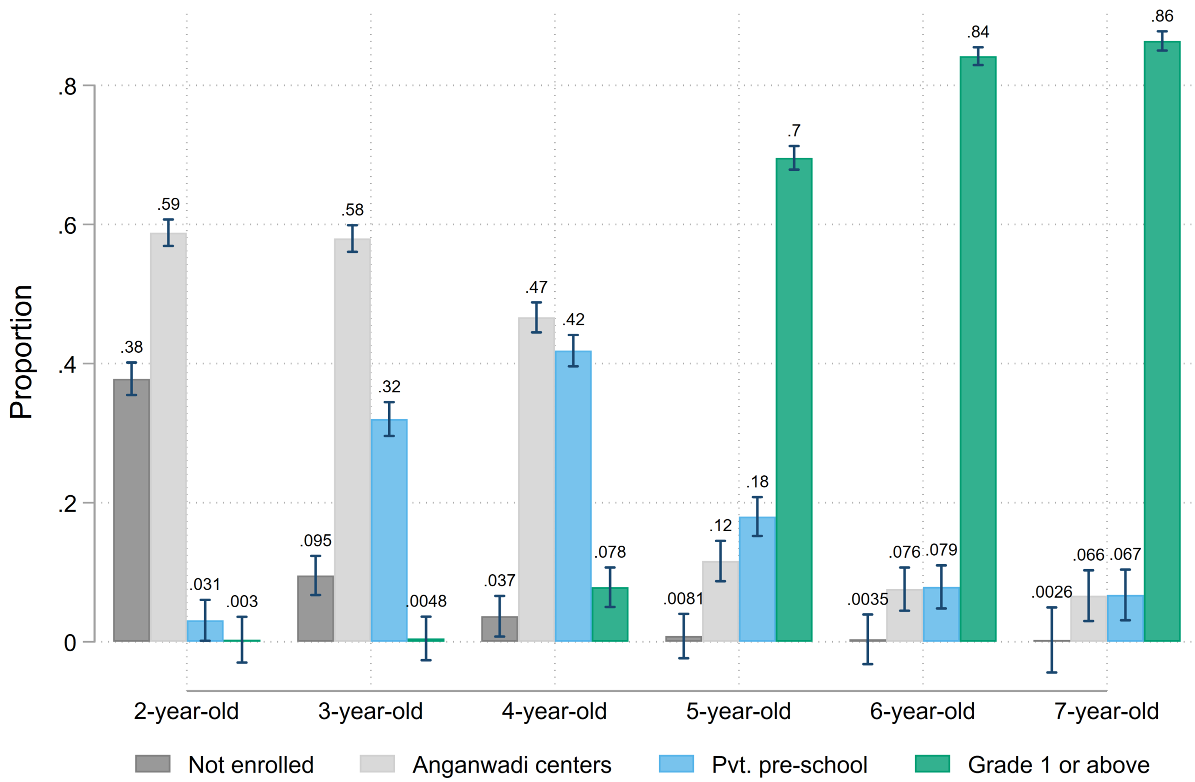
a NSS Report No. 554 (July 2011-June 2012), statement 4.1.2.
b NSS Report No. 554 (July 2011-June 2012), statement 5.13.1.
c MOSPI states that real gross state domestic product in Tamil Nadu grew 6.4% per year from 2011 to 2017.
d Table 5, Panel A, column 4.
e Preferred estimate from Kline and Walters (2016), Appendix Table A.IV.
f Table 6, bottom panel, columns 1 and 2.
g Hoddinott et al. (2011) estimate that a standard deviation increase in HAZ increases adult consumption by 20% and that stunting reduces adult consumption by 66%.
h Table A.10, Panel B, column 4. We adjust for this association by subtracting the implied effect of nutrition on test scores from the total test score effect before projecting earnings gains.
i Assuming 25% turnover per year, the experimental data understate the number of children served over two years by 33%.

Figure A.1: Tamil Nadu in India and sampled districts



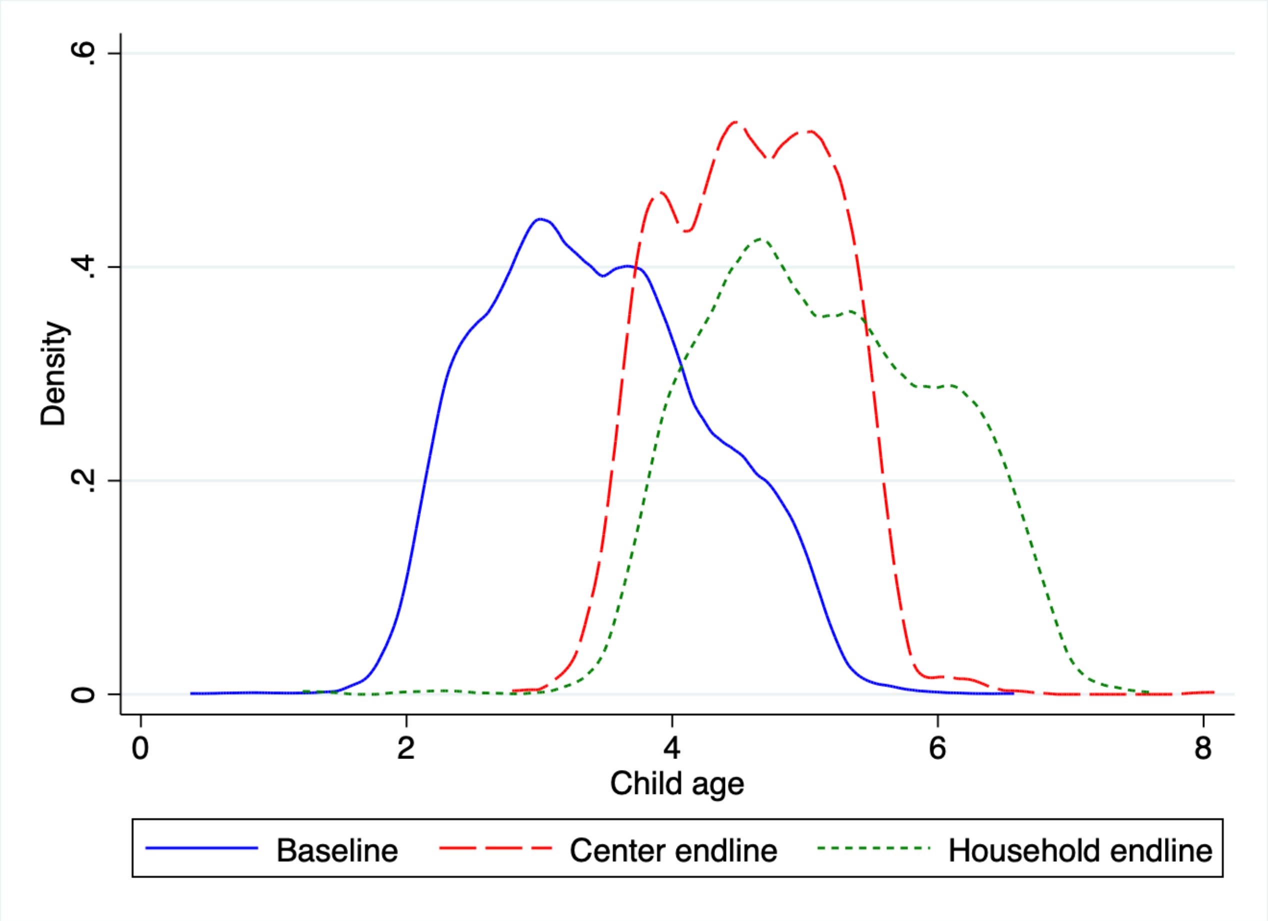
Notes: The figure shows the state of Tamil Nadu in India (on the left) and the four sampled districts (shaded): Kancheepuram, Karur, Trichy, and Virudhunagar. Sampling was stratified by geographic zone and average nutrition status.

Figure A.2: Proportion of children by age and enrollment in pre-primary education



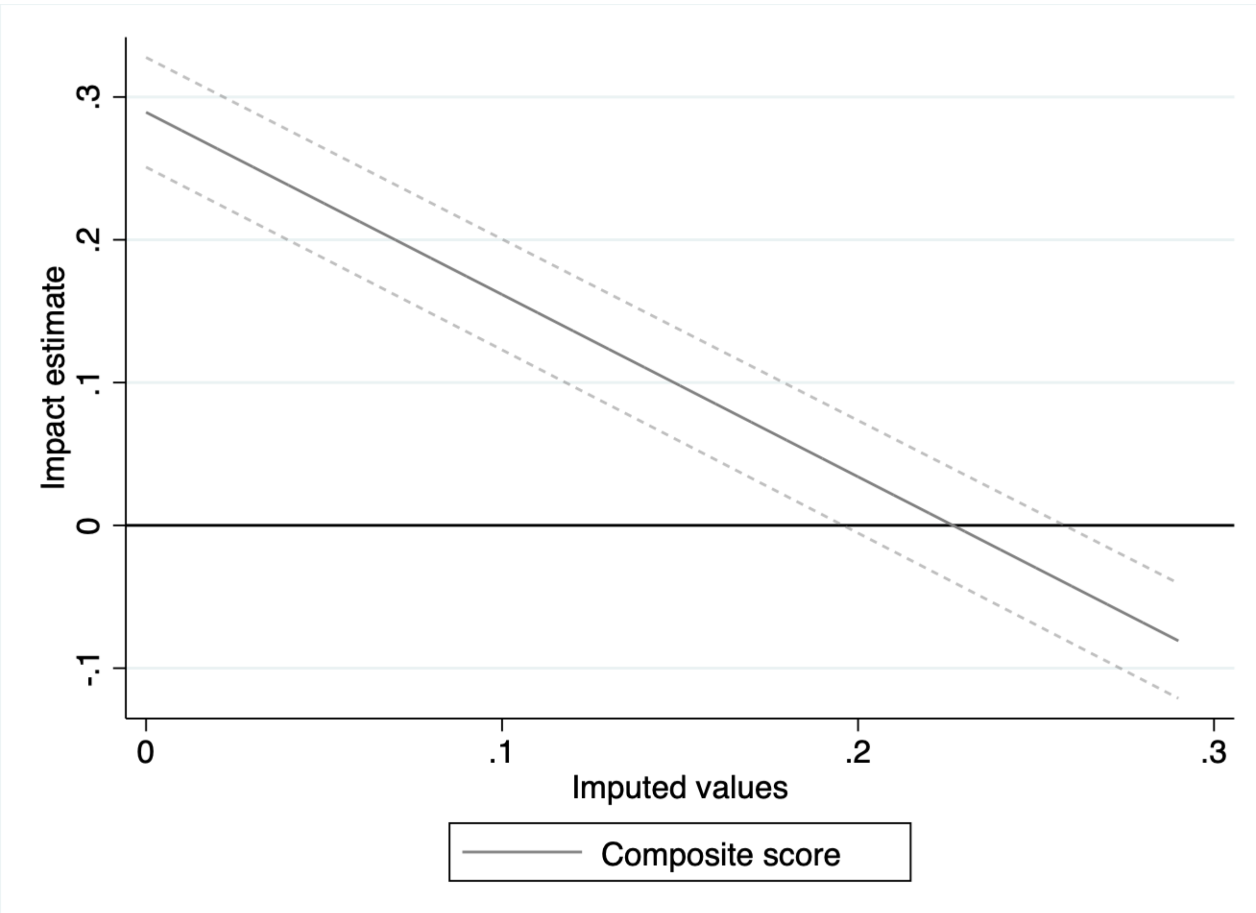
Notes: The figure shows the share of children by age and enrollment status using a representative household survey in a different set of villages in the same study districts. This data is from a different (ongoing) project studying household choice behavior across preschooling options and the sample comprises 23,717 children (aged 2-7) across 197 villages.

Figure A.3: Children's age distribution by round of data collection



Notes: The figure shows the distribution of children's ages at each round of data collection: the baseline (blue solid line) and endline (red dashed line for AWC measurements and green dashed line for HH measurements). The figure includes all children in the estimation samples for a given round. The distribution for the HH endline is weighted by the inverse sampling probability for the HH survey.

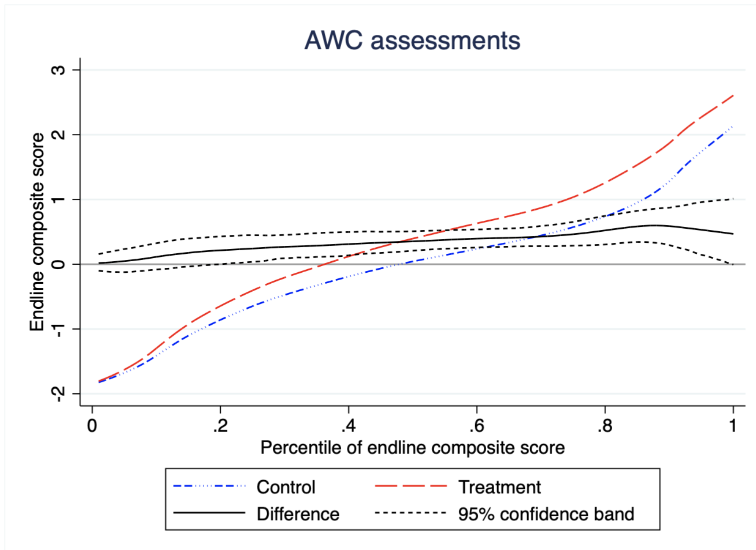
Figure A.4: AWC composite score impact estimates with outcomes imputed for missing children



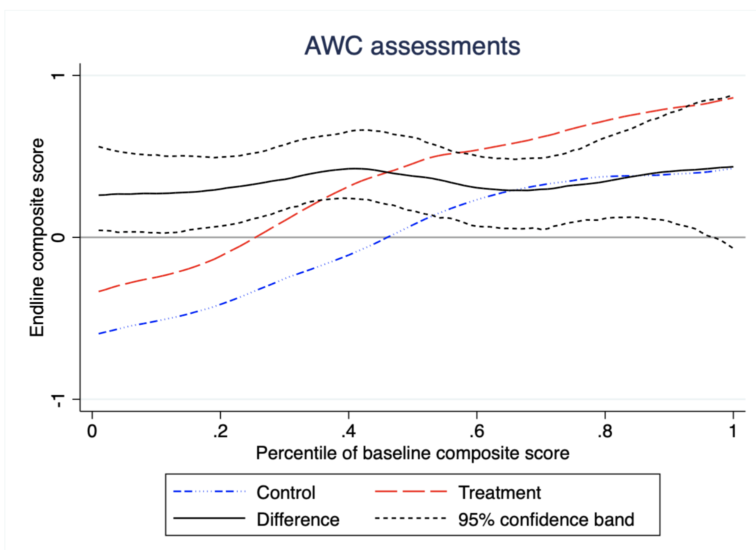
Notes: The figure shows estimated effects on standardized endline composite scores in the AWC sample with missing outcomes imputed for attriting children. The sample includes all 4,521 children observed at baseline with non-missing baseline covariates. Missing outcomes for control children are imputed as the control mean plus the control standard deviation times the value on the horizontal axis. Missing outcomes for treated children are imputed as the treated mean minus the treated standard deviation times the value on the horizontal axis. Effects are estimated using the same specification from Table 5.

Figure A.5: Distributional treatment effects on achievement

A. Quantile treatment effects



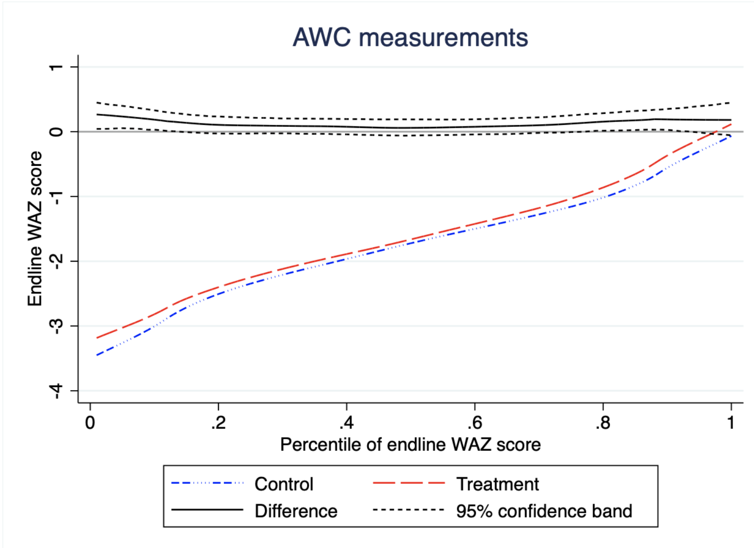
B. Average treatment effects by baseline score



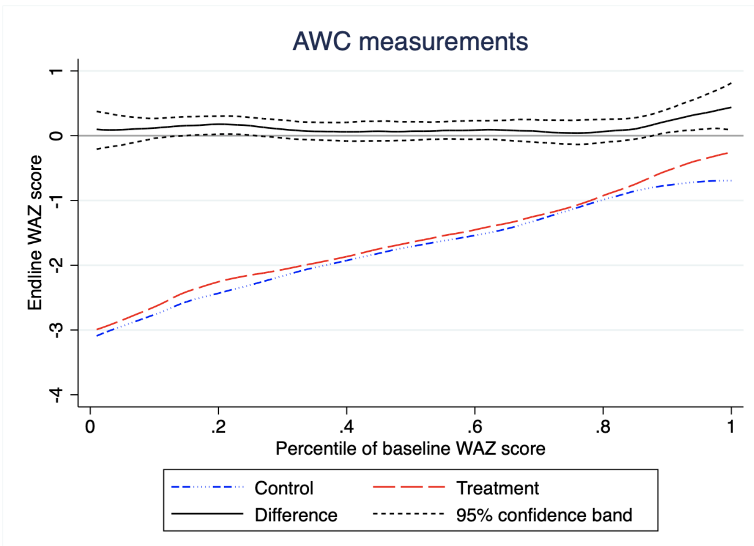
Notes: Panel A shows quantiles of endline composite test scores for treated and control children who participated in the baseline and endline AWC assessments, estimated by local polynomial regressions of endline scores on endline percentiles separately by experimental group. The solid black line plots the difference between treatment and control (quantile treatment effects). Panel B shows estimates of average endline composite scores and treatment effects at each percentile of baseline composite score, estimated by local polynomial regression. Dashed black lines display bootstrapped 95% confidence intervals.

Figure A.6: Distributional treatment effects on WAZ scores

A. Quantile treatment effects



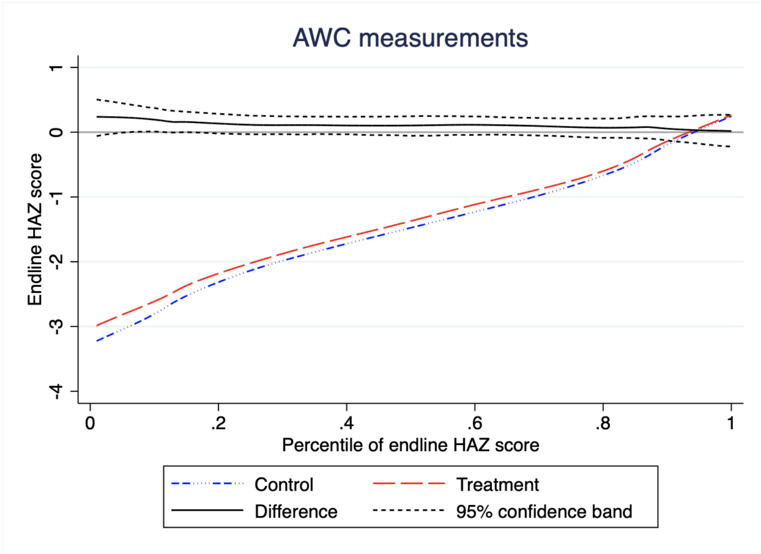
B. Average treatment effects by baseline score



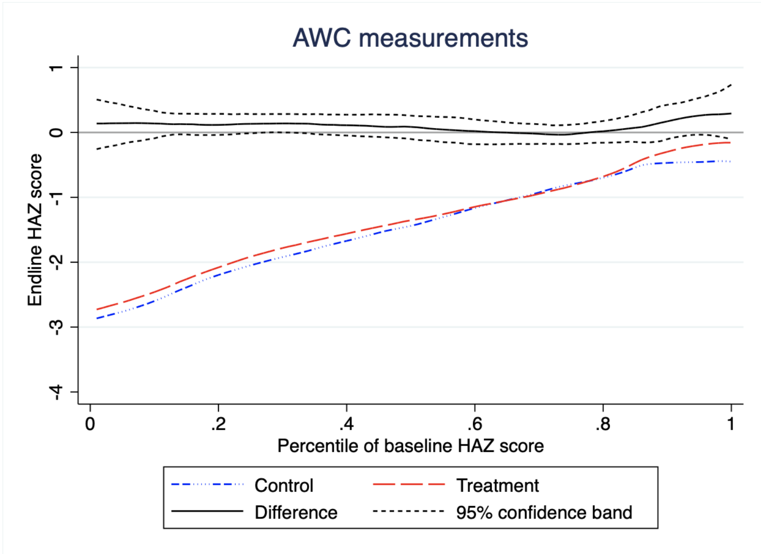
Notes: Panel A shows quantiles of endline weight-for-age z-scores (WAZ) for treated and control children who participated in the baseline and endline AWC assessments, estimated by local polynomial regressions of endline scores on endline percentiles separately by experimental group. The solid black line plots the difference between treatment and control (quantile treatment effects). Panel B shows estimates of average endline WAZ scores and treatment effects at each percentile of baseline WAZ score, estimated by local polynomial regression. Dashed black lines display bootstrapped 95% confidence intervals.

Figure A.7: Distributional treatment effects on HAZ scores

A. Quantile treatment effects

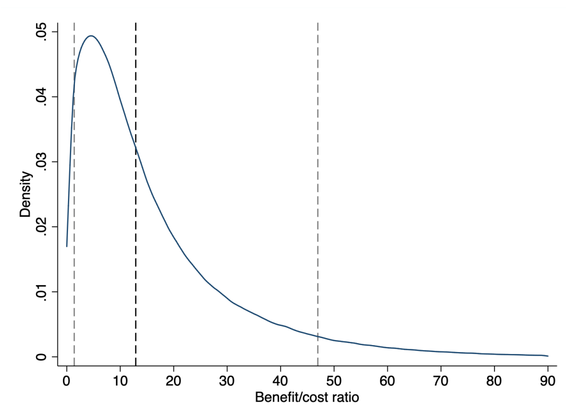


B. Average treatment effects by baseline score

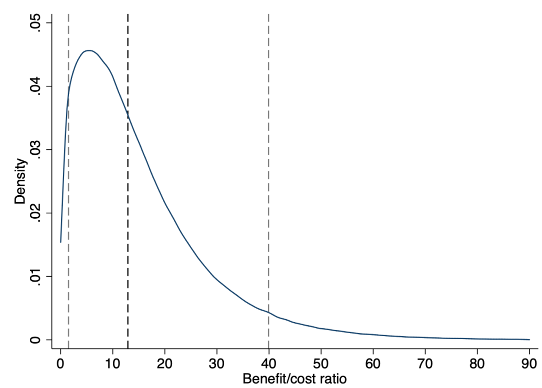


Notes: Panel A shows quantiles of endline height-for-age z-scores (HAZ) for treated and control children who participated in the baseline and endline AWC assessments, estimated by local polynomial regressions of endline scores on endline percentiles separately by experimental group. The solid black line plots the difference between treatment and control (quantile treatment effects). Panel B shows estimates of average endline HAZ scores and treatment effects at each percentile of baseline HAZ score, estimated by local polynomial regression. Dashed black lines display bootstrapped 95% confidence intervals.

Figure A.8: Benefit-cost sensitivity analysis (log-normal discount rate)



Uniform parameters (mean = 16.1, median = 11.3)



Truncated normal parameters (mean = 15.3, median = 11.9)

Notes: This figure explores the sensitivity of the ECCE facilitator benefit/cost ratio to parameter values. Estimates are based on test-score gains for the household sample as in column (1) of Table 5. We obtain a distribution of benefit/cost ratios by drawing each parameter calibrated from other data sources from a range of possible values, with the preferred values from Table 5 in the middle of each range. Days worked in the labor market range from 200 to 250. The wage growth rate ranges from 3 percent to 7 percent. The proportionate increase in earnings per standard deviation of test scores ranges from 7 percent to 19 percent. The left-hand plot draws each of these parameters from an independent uniform distribution, while the right-hand plot draws each parameter from an independent truncated normal distribution with mean in the middle of the range and standard deviation 1/4 of the width of the range. In both panels the discount rate is drawn from a log-normal distribution with a median of 3 percent and a standard deviation of 3 percent. Results come from fitting kernel densities to 1,000,000 draws of the parameters, excluding values over 90 for visual presentation. Gray lines indicate 5th and 95th percentiles, and black vertical lines denote our preferred estimates from Table 7.

Appendix B Supplemental robustness checks

Table B.1: Comparison of in- and out-of-sample districts

	(1) Out-of-sample	(2) In-sample	(3) Difference	(4) N
Proportion of children with normal weight	0.879	0.896	0.017	31
	[0.051]	[0.067]	(0.028)	
Proportion of moderately underweight children	0.120	0.103	-0.017	31
	[0.051]	[0.066]	(0.028)	
Proportion of severely underweight chidlren	0.001	0.001	-0.000	31
	[0.001]	[0.001]	(0.000)	
Proportion of underweight children	0.001	0.000	-0.001	31
	[0.001]	[0.000]	(0.001)	
Number of residents in the district (in 1000s)	2,139.740	2,431.831	292.091	31
	[957.586]	[1,244.629]	(531.013)	

Notes: This table compares the districts that were selected and those that were not selected for the sample on the variables used for sampling. Standard deviations appear in brackets. Note that the random sampling of districts was conducted before our baseline (whereas random assignment of AWCs to treatment and control status was done after the baseline). This table therefore compares sampled districts and out-of-sample districts using administrative data on under-nutrition. As has been shown in other settings, administrative data significantly overstates outcomes and understates problems. Thus, this table is only meant to show that the representative nature of the sampled districts.

Table B.2: Facilitator self-reports on time allocation from intervention monitoring

Hours spent last week...	(1)
A. Education	
...teaching pre-school education	17.452
	[9.198]
B. Health and nutrition	
...providing nutrition to children	0.360
	[1.338]
...providing nutrition to mothers	0.177
	[0.821]
...cooking/distributing food	0.399
	[1.440]
...measuring weight/height of children	0.155
	[0.492]
...conducting health check-ups	0.116
	[0.744]
...conducting home visitations	0.262
	[0.987]
Sub-total	1.469
	[3.458]
C. Administrative	
...cleaning	0.524
	[2.542]
...recruiting beneficiaries	0.013
	[0.144]
...picking up children to come to the AWC	0.539
	[1.351]
...organizing awareness campaigns	0.231
	[1.370]
...preparing informational materials	2.264
	[4.396]
...preparing monthly progress reports	0.031
	[0.207]
...maintaining ECE registers	1.210
	[2.819]
...maintaining other registers	0.329
	[1.222]
Sub-total	5.141
	[6.291]
Total	24.062
	[11.542]
N (centers)	160

Notes: This table shows descriptive statistics for facilitators in treatment centers for all variables collected in the first round of intervention monitoring, from April to May of 2017. Only the variable on salary was collected in the second round. All intervention-monitoring visits were pre-scheduled. Standard deviations appear in brackets. In reporting 17.5 hours per week for preschool education, it seems likely that facilitators are considering preparation time and other miscellaneous additional time spent with children in the center (e.g., assisting with nap time).

Table B.3: Impact on endline assessments (proportion-correct scores)

	(1) Math	(2) Language	(3) Executive function	(4) Composite score
A. Complete sample				
AWC assessments (N=1514)				
Treatment	0.052***	0.074***	0.055***	0.060***
	(0.011)	(0.013)	(0.016)	(0.012)
Control mean	0.290	0.204	0.566	0.376
HH assessments (N=2075)				
Treatment	0.041***	0.033**	0.015	0.027**
	(0.016)	(0.017)	(0.011)	(0.012)
Control mean	0.465	0.455	0.503	0.477
B. Common sample				
AWC assessments (N=791)				
Treatment	0.055***	0.074***	0.062***	0.063***
	(0.013)	(0.015)	(0.021)	(0.014)
Control mean	0.287	0.199	0.567	0.374
HH assessments (N=791)				
Treatment	0.061***	0.077***	0.039**	0.057***
	(0.017)	(0.019)	(0.015)	(0.014)
Control mean	0.281	0.263	0.438	0.336
P-value (AWC = HH)	0.621	0.763	0.207	0.602

Notes: The table shows the impact of the intervention on assessments of math, language, and executive function after two years. Estimates come from regressions of endline test scores on a treatment indicator with controls for randomization strata and baseline characteristics. All scores are expressed as proportions of questions in each test answered correctly. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline assessments. Panel B displays results for all children who participated in the baseline and both endline assessments. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. The last row displays the p-value testing the null hypothesis that the treatment effects across both assessments in Panel B are equal. Standard errors (clustered by AWC) appear in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.4: Impact on endline assessments (any-correct answers)

	(1) Math	(2) Language	(3) Executive function	(4) Composite score
A. Complete sample				
AWC assessments (N=1514)				
Treatment	-0.006	0.035*	0.009	0.005
	(0.013)	(0.019)	(0.012)	(0.012)
Control mean	0.933	0.837	0.939	0.947
HH assessments				
Treatment	0.038**	0.040**	0.040***	0.040***
	(0.015)	(0.016)	(0.013)	(0.013)
Control mean	0.871	0.856	0.898	0.902
B. Common sample				
AWC assessments (N=791)				
Treatment	0.002	0.056**	0.016	0.017
	(0.019)	(0.028)	(0.018)	(0.017)
Control mean	0.937	0.837	0.944	0.949
HH assessments (N=791)				
Treatment	0.067***	0.067**	0.056**	0.055**
	(0.026)	(0.028)	(0.022)	(0.022)
Control mean	0.832	0.798	0.888	0.890

Notes: The table shows the impact of the intervention on assessments of math, language, and executive function after two years. Estimates come from regressions of indicator variables for any correct answers on a treatment indicator with controls for randomization strata and baseline characteristics. Panel A displays results for all children who participated in the baseline assessment, separately for children in the AWC and household (HH) endline assessments. Panel B displays results for all children who participated in the baseline and both endline assessments. Estimates for the full HH sample weight by the inverse sampling probability for the HH survey. Estimates for the AWC and common samples do not use weights. All specifications control for a baseline measure of the dependent variable, AWW experience, and AWW education. Standard errors (clustered by AWC) appear in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.5: Impact on endline assessments on the same sample and items

	Math		Language		Executive function		Composite score
	(1) Raw	(2) Std.	(3) Raw	(4) Std.	(5) Raw	(6) Std.	(7) Std.
A. AWC (N=791)							
Treatment	0.078***	0.385***	0.078***	0.399***	0.074***	0.256***	0.388***
	(0.016)	(0.082)	(0.018)	(0.091)	(0.023)	(0.080)	(0.082)
Control mean	0.221	-0.000	0.189	0.000	0.560	-0.000	-0.000
B. Household (N=791)							
Treatment	0.061***	0.272***	0.077***	0.328***	0.039**	0.171**	0.319***
	(0.017)	(0.076)	(0.019)	(0.082)	(0.015)	(0.067)	(0.078)
Control mean	0.281	0.000	0.263	-0.000	0.438	0.000	0.103
P-value (T $\times$ AWC)	0.576	0.362	0.635	0.213	0.267	0.503	0.344

Table B.6: Impacts on WAZ scores after dropping child outliers based on residuals

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
Dropping residuals > 0.5 (N=1131)			
Treatment	0.086**	-0.023	-0.030***
	(0.034)	(0.021)	(0.010)
Control mean	-1.898	0.427	0.103
Dropping residuals > 0.4 (N=1043)			
Treatment	0.083**	-0.026	-0.033***
	(0.035)	(0.021)	(0.011)
Control mean	-1.940	0.444	0.108
Dropping residuals > 0.3 (N=959)			
Treatment	0.106***	-0.037	-0.038***
	(0.035)	(0.023)	(0.012)
Control mean	-2.004	0.472	0.120
Dropping residuals > 0.2 (N=865)			
Treatment	0.109***	-0.052**	-0.042***
	(0.038)	(0.025)	(0.014)
Control mean	-2.076	0.512	0.133
Dropping residuals > 0.1 (N=752)			
Treatment	0.122***	-0.068**	-0.049***
	(0.043)	(0.029)	(0.016)
Control mean	-2.186	0.568	0.155

Notes: The table replicates the estimation from panel A of Table 6 with different levels of residual outlier exclusions. We first regress WAZ scores on a treatment indicator and baseline WAZ scores, and form residuals from this regression. A child is marked as an outlier if the absolute value of the residual from this model is greater than the relevant threshold. We then estimate the impact of the treatment excluding these outliers. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.7: Impacts on HAZ scores after dropping child outliers based on residuals

	(1) HAZ score	(2) Stunted (HAZ<-2)	(3) Severely stunted (HAZ<-3)
Dropping residuals > 0.5 (N=1131)			
Treatment	0.122***	-0.058**	-0.021*
	(0.044)	(0.026)	(0.011)
Control mean	-1.782	0.361	0.070
Dropping residuals > 0.4 (N=1043)			
Treatment	0.148***	-0.078***	-0.025**
	(0.043)	(0.028)	(0.012)
Control mean	-1.867	0.394	0.077
Dropping residuals > 0.3 (N=959)			
Treatment	0.162***	-0.091***	-0.028**
	(0.044)	(0.030)	(0.013)
Control mean	-1.950	0.432	0.084
Dropping residuals > 0.2 (N=865)			
Treatment	0.172***	-0.097***	-0.031**
	(0.047)	(0.033)	(0.014)
Control mean	-2.038	0.477	0.094
Dropping residuals > 0.1 (N=752)			
Treatment	0.176***	-0.118***	-0.039**
	(0.053)	(0.038)	(0.017)
Control mean	-2.131	0.540	0.107

Notes: The table replicates the estimation from panel B of Table 6 with different levels of residual outlier exclusions. We first regress HAZ scores on a treatment indicator and baseline HAZ scores, and form residuals from this regression. A child is marked as an outlier if the absolute value of the residual from this model is greater than the relevant threshold. We then estimate the impact of the treatment excluding these outliers. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.8: Impacts on winsorized WAZ scores

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
Winsorized at 0.1% (N=1538)			
Treatment	0.104***	-0.021	-0.025***
	(0.034)	(0.018)	(0.009)
Control mean	-1.757	0.384	0.091
Winsorized at 0.2% (N=1538)			
Treatment	0.107***	-0.021	-0.025***
	(0.035)	(0.018)	(0.009)
Control mean	-1.759	0.384	0.091
Winsorized at 0.5% (N=1538)			
Treatment	0.106***	-0.021	-0.025***
	(0.036)	(0.018)	(0.009)
Control mean	-1.760	0.384	0.091
Winsorized at 1% (N=1538)			
Treatment	0.105***	-0.021	-0.025***
	(0.036)	(0.018)	(0.009)
Control mean	-1.761	0.384	0.091
Winsorized at 2% (N=1538)			
Treatment	0.105***	-0.021	-0.025***
	(0.037)	(0.018)	(0.009)
Control mean	-1.762	0.384	0.091

Notes: The table replicates the estimation from panel A of Table 6 with different levels of winsorizing. The endline outcome measure is winsorized based on the given value. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.9: Impacts on winsorized HAZ scores

	(1) HAZ score	(2) Stunted (HAZ<-2)	(3) Severely stunted (HAZ<-3)
Winsorized at 0.1% (N=1538)			
Treatment	0.077*	-0.041**	-0.016*
	(0.045)	(0.020)	(0.008)
Control mean	-1.477	0.291	0.057
Winsorized at 0.2% (N=1538)			
Treatment	0.080*	-0.041**	-0.016*
	(0.046)	(0.020)	(0.008)
Control mean	-1.482	0.291	0.057
Winsorized at 0.5% (N=1538)			
Treatment	0.084*	-0.041**	-0.016*
	(0.047)	(0.020)	(0.008)
Control mean	-1.487	0.291	0.057
Winsorized at 1% (N=1538)			
Treatment	0.084*	-0.041**	-0.016*
	(0.048)	(0.020)	(0.008)
Control mean	-1.490	0.291	0.057
Winsorized at 2% (N=1538)			
Treatment	0.084*	-0.041**	-0.016*
	(0.048)	(0.020)	(0.008)
Control mean	-1.490	0.291	0.057

Notes: The table replicates the estimation from panel B of Table 6 with different levels of winsorizing. The endline outcome measure is winsorized based on the given value. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix C Measurement

C.1 Child assessments

C.1.1 Test construction

The assessments of math, language, and executive function skills were designed by the research team, drawing on assessments with evidence of validity and reliability in developing countries (see, for example, Araujo et al., 2016; Wolf et al., 2017; Obradović et al., 2014; Halpin et al., 2019). They were administered individually, orally, and in Tamil by local enumerators hired, trained, and monitored by J-PAL South Asia.

The baseline assessments included few items because their main purpose was to allow us to account for children's initial learning levels in our impact estimations. The math test asked children to count, collect sets of toys involving different quantities, and identify numbers. The language test asked children to identify letters. The executive function test included shape and color card sorts and a Stroop white-black test to measure children's cognitive flexibility and a digit span and an ordered-object recognition task to measure their short-term memory.[40]

The endline assessments included additional items to allow us to estimate the impact of the intervention with greater precision. The math test asked children to compare drawings based on their shape, length, and quantity. The language test asked them to name foods and animals to measure their expressive vocabulary, answer questions on a short story to measure their listening comprehension, and manipulate a storybook to demonstrate their print awareness. The executive function test included two games to measure children's inhibitory control. The assessments administered in households included a subset of the items administered in *anganwadi* centers: 12 of the 24 items in math, 15 of the 20 items in language, and 17 of the 29 items in executive function.[41]

C.1.2 Test-score distributions

We calculated each child's score on each subject as proportion-correct scores, both raw and standardized with respect to the overall baseline distribution. The mean raw scores by experimental group are shown in Table C.1. The mean standardized scores are in Table 1.

Figure C.1 displays the distribution of raw scores for each round of administration of the assessments. As the graph shows, the baseline assessments on math and language were too difficult for many children in our study: 49% of children could not answer any questions in math and 69% could not answer any questions in language (there were no statistically significant differences across experimental groups). This was not the case in executive function,

[40]The baseline assessments can be accessed at: `https://bit.ly/2KGITA9`.
[41]The endline assessments can be accessed at: `https://bit.ly/2P900eP`.

for which only 7% of children. This pattern may be attributed in part to the enrollment of many young children in our study centers: nearly 33% of children taking the baseline assessments were below the age of 3 (Figure A.3).

The endline assessments were more appropriate for our study sample. Only 25% of children who took the assessment at the center could not answer any questions in math, 38% in language, and 22% in executive function. The corresponding figures for the children who took the assessments at their homes were 21%, 26%, and 16%, respectively (Figure C.1).

C.2 Visits to *anganwadi* centers

The unannounced visits to AWCs measured the effect of the intervention on worker attendance and punctuality and overall time allocation. The announced visits measured the effect of the intervention on instructional time use. Enumerators arrived at each center before the official start of preschool education and tracked the amount of time that the worker and facilitator spent on instructional activities for the two hours devoted to preschool education, using another adaptation of the Stallings protocol. We conducted these observations in a random sample of 20 AWCs per district (i.e., 10 per experimental arm), for a total of 80 centers.[42]

C.3 Children's weight and height

Enumerators from J-PAL South Asia measured each child's weight as follows. First, they removed the child's shoes, headpieces, accessories, and jewelry; they checked that the child's pockets were empty; and made sure that the child was wearing light-weight clothes and bare feet. Then, they placed the scale on a hard and flat surface and made sure that the child was standing in the center of the scale, looking straight ahead, and with the weight evenly divided on both feet. Once these two steps were completed, the enumerator and worker repeated the process and recorded the weight a second time.

Enumerators also measured each child's height as follows. First, they assembled the stadiometer and placed it against a wall, ensuring it was stable. Then, they removed the child's shoes; pushed aside any hair that would interfere with the height measurement; and made sure that the child was standing on the base of the stadiometer and facing forward. They placed the child's feet flat and together in the center of the base, checking that the child's legs were straight, his/her buttocks were touching the stadiometer, his/her shoulders are even, and his/her hands are on the sides. Finally, the child was asked to take a deep

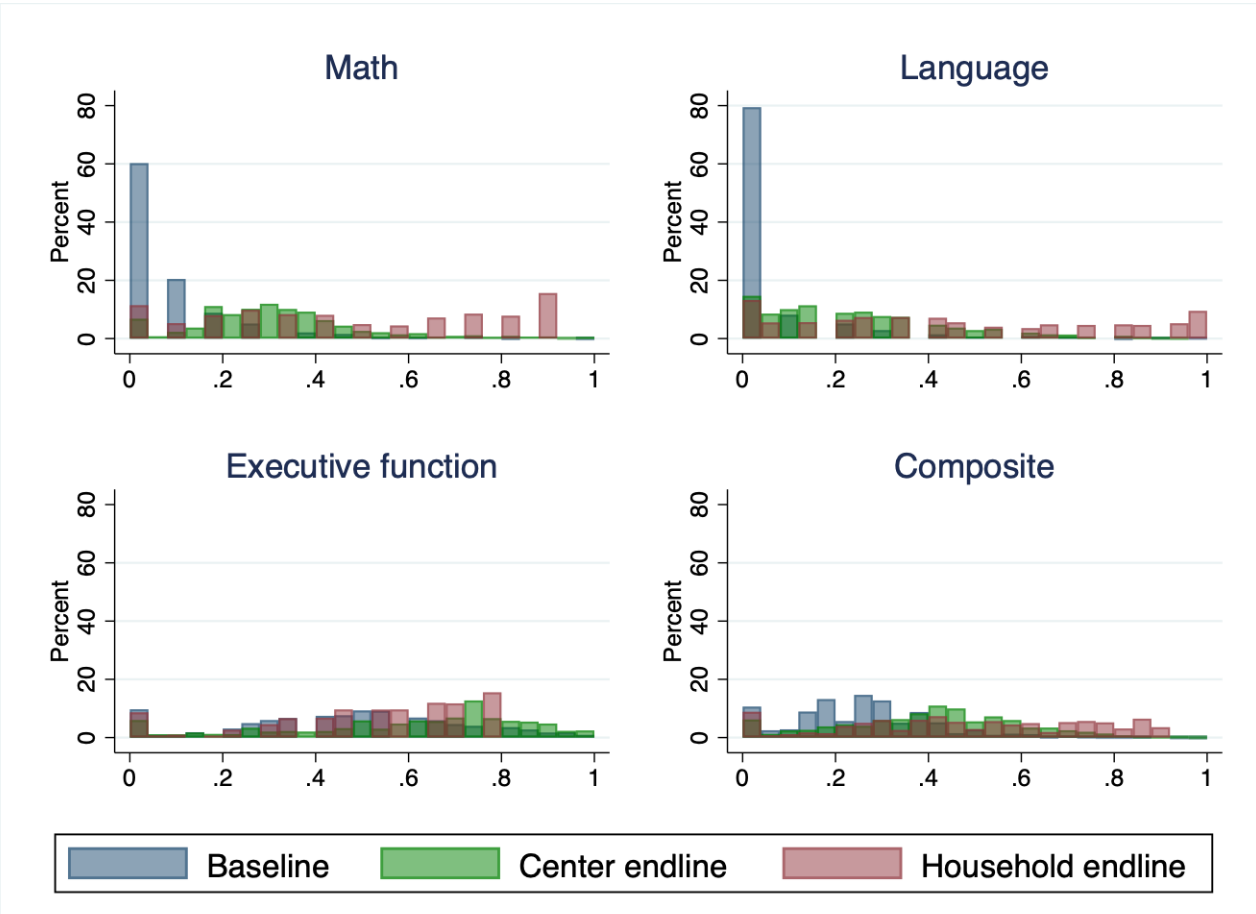
[42]The protocols for the announced visits can be accessed at: `https://bit.ly/2DZri56`.

breath and hold it while his/her height was measured. Once these steps were completed, the enumerator and worker repeated the process and recorded the height a second time.[43]

To assess the reliability of our measurement protocol, 15 percent of AWC observations and 50 percent of HH observations were randomly selected for repeat back-check measurements within 2-4 days of the main measurement. For measurements taken at the AWC, correlations between the main and back-check measurements were 0.959 for weight and 0.964 for height. Corresponding correlations for the HH measurements were 0.807 for weight and 0.757 for height.

[43] The protocols for the measurements can be accessed at: `https://bit.ly/2VtT1lF`.

Figure C.1: Distribution of proportion-correct scores in assessments by round of data collection



Notes: The figure shows the distribution of the proportion-correct scores on the math, language, and executive function assessments for children in the estimation samples. Each proportion-correct score indicates the proportion of items answered correctly in a subject. The figure includes children with a baseline score and with either a center or a household score.

Table C.1: Raw proportion-correct scores at baseline

	(1) Control	(2) Treatment	(3) Difference	(4) N
Math (proportion-correct score)	0.117	0.125	0.003	4,675
	[0.179]	[0.181]	(0.008)	
Language (proportion-correct score)	0.093	0.097	-0.002	4,675
	[0.191]	[0.191]	(0.009)	
Exec. function (proportion-correct score)	0.535	0.537	-0.002	4,675
	[0.249]	[0.257]	(0.011)	

Notes: This table compares children's learning outcomes in the control and treatment groups at baseline. It shows the means and standard deviations for each group (columns 1-2) and tests for differences between groups including randomization-strata fixed effects (column 3). The sample includes all children observed at baseline. Standard deviations appear in brackets, and standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

References

Araujo, M. C., P. M Carneiro, Y. Cruz-Aguayo, and N. Schady (2016). “Teacher Quality and Learning Outcomes in Kindergarten.” *The Quarterly Journal of Economics* 131 (3), pp. 1415–1453.

Halpin, P., S. Wolf, H. Yoshikawa, S. Kabay, and N. Rojas (2019). “Measurement invariance of the International Development and Early Learning Assessment (IDELA) across five countries.” *Developmental Psychology* 55 (1), pp. 23–37.

Obradović, J., A. K. Yousafzai, X. A. Portilla, M. A. Rasheed, U. Memon, S. Siyal, N. Tirado-Strayer, and B. Keehn (2014). “School Readiness and Transition (STAR) study: Executive functions battery adaptation manual. Early Childhood Cognitive Stimulation and Successful Transition to Preschool in a Disadvantaged Population in Rural Pakistan.” *Unpublished manuscript*. Harvard University.

Wolf, S., P. Halpin, H. Yoshikawa, A. J. Dowd, L. Pisani, and I. Borisova (2017). “Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia.” *Early Childhood Research Quarterly* 41, pp. 21–36.

Appendix D Benefit-Cost Calculations

D.1 Baseline benefit-cost analysis

We assess cost-effectiveness of the AWC facilitator intervention by comparing a prediction of the present discounted value of earnings gains generated by the program to the intervention's cost. The benefit-cost ratio is calculated as:

$$R = \frac{\left(\sum_{a=A_{min}}^{A_{max}} PDV(a)\right) \times \Delta Y \times P \times N}{C},$$

where $PDV(a)$ is the average present discounted value of children's earnings at age a, ΔY is the test score gain caused by the program, P is the labor market value of test score gains (i.e. a skill price), N is the number of children treated by the program, and C is the program's cost.

Children are assumed to work from ages $A_{min} = 22$ to $A_{max} = 65$. The present value of earnings at each age is given by

$$PDV(a) = \frac{\bar{W} \times D \times LFP \times \times (1+g)^{a+4-5}}{(1+r)^{a-5}},$$

where $\bar{W}$ is the current average daily wage for workers in rural Tamil Nadu, D is the number of days worked per year when in the labor force, LFP is the labor force participation rate, g is the annual real wage growth rate, and r is the discount rate. The -5 in the exponents in the numerator and denominator reflects the fact that we are discounting costs and benefits back to the end of a child's age 4 year, while the extra 4 in the exponent of the numerator reflects the fact that if the program is implemented in steady state we expect an extra 4 years of wage growth to accrue between birth and the time that program costs are incurred for each cohort of children. The term in parentheses then measures the present discounted value of a child's age a earnings at the time he or she enrolls in the AWC. Given the parameter calibrations in Table 7, we calculate the total PDV of earnings for a child in rural Tamil Nadu to equal $\sum_{a=A_{min}}^{A_{max}} PDV(a) =$ INR 3.62 million.

We measure the test score gain ΔY using the estimated causal impact of the intervention on composite test scores for the HH sample, which covers the full initial treated cohort of children including those that later attrited from the AWC. As shown in column 4 of the second row of Table 5, this estimate is 0.11 standard deviations. The parameter P measures the predicted proportionate increase in earnings associated with a one standard deviation increase in test scores. Based on estimates from Chetty et al. (2011) as well as a review of such estimates in Kline and Walters (2016), we calibrate this parameter to equal 0.13. As a result, the total predicted increase in earnings for a given child is roughly INR 3.62 million $\times 0.11 \times 0.13 =$ INR 52,000.

The program treated 14 children per center in the baseline cohort as well as an additional cohort not measured at baseline (and therefore not included in the analysis); with roughly 25% turnover across cohorts this implies that the experimental sample understates the total number of treated children by 33%, so we set $N = 14 \times 1.33$. This yields a total predicted benefit per center of INR 52,000 $\times 14 \times 1.33 =$ INR 964,000. The cost of the program was about $C =$ INR 74,000. Taking the ratio of these benefits and costs yields an estimated R of 12.9 in our baseline calibration.

D.2 Sensitivity analysis

This cost effectiveness calculation depends on several parameters calibrated from external sources. To assess the sensitivity of our results to these choices, we compute a distribution of benefit-cost ratios using parameters drawn from a wide set of possible values. The left-hand panel of Figure 1 draws parameters from the following independent uniform distributions, centered at the preferred values from Table 7:

$$\begin{aligned} D &\sim Uniform(200, 250), \\ g &\sim Uniform(0.03, 0.07), \\ r &\sim Uniform(0.015, 0.045), \\ P &\sim Uniform(0.07, 0.19). \end{aligned}$$

The right-hand panel of Figure 1 draws parameters from truncated normal distributions with support on these same values and standard deviations equal to one-fourth of the width of the support for each parameter. The results reveal large benefit-cost ratios for most parameters we consider, with 5th percentiles of $R = 4.2$ and $R = 5.5$ in the two panels, and 5 percent of values in excess of $R = 30$ in each panel.

Since any earnings gains accrue many years into the future, benefit-cost calculations for early-childhood programs may be especially sensitive to the assumed discount rate. Practitioners adopt a variety of approaches to choosing discount rates, and some methods are likely to yield higher values than the 3 percent value assumed in our baseline analysis, which is based on the real return on government bonds (Dhaliwal et al., 2013). To allow for a wider range of discount rates and a thicker upper tail with more high values, Figure A.8 repeats the sensitivity analysis by drawing r from a log-normal distribution with location parameter $\mu = \log(0.03)$ and scale parameter $\sigma^2 = \log\left(\frac{1+\sqrt{5}}{2}\right)$, which results in a distribution with median 3 percent and standard deviation 3 percent. These distributions turn out to be similar to those in Figure 1, though the thicker upper tail for r yields a few more small values of the benefit-cost ratio, with 5th percentiles equal to 1.2 and 1.3 in the left and right panels, respectively.

D.3 Accounting for nutrition gains

Our baseline cost-effectiveness calculation ignores nutrition benefits because estimates of nutrition effects are statistically insignificant in the full HH sample. Since we see improvements in nutrition in the AWC sample, however, it's worth asking how incorporating these impacts changes the benefit-cost ratio ratio R.

Appendix Table A.13 assesses this issue by reporting benefit-cost ratios based on the AWC sample. Column 2 continues to ignore nutrition benefits and computes R based on test score gains in the AWC sample. This calculation replaces ΔY with the larger value of 0.29 standard deviation reported in the first row of column 4 in Table 5, and reduces the number of children per center from 14 to 5 since on average only 5 members of the baseline cohort appeared at the AWC at endline. This results in only a small change to the benefit-cost ratio, reducing R from 12.9 to 12.2. The similarity of these values is a consequence of the fact that the overall test score impact in the HH sample is roughly equal to the effect in the AWC sample times the participation rate, as discussed in Section 4.

Since nutrition gains are significant in the AWC sample, we extend the benefit-cost calculation for this group to incorporate nutrition benefits. The benefit-cost formula in this case equals:

$$R = \frac{\left(\sum_{a=A_{min}}^{A_{max}} PDV(a)\right) \times [(\Delta Y_t - \Delta Y_n \gamma_{tn}) \times P_t + \Delta Y_n \times P_n] \times N}{C},$$

where ΔY_t and ΔY_n now measure the intervention's impacts on test scores and nutrition, and P_t and P_n measure the earnings gains associated with improvements in each outcome. The term γ_{tn} measures the effect of improved nutrition on test scores; we subtract $\Delta Y_n \gamma_{tn}$ from the test score gain before applying the skill price P_t to avoid double-counting test score gains that accrue through the nutrition channel.

Column 3 reports a benefit-cost calculation that measures the nutrition gain ΔY_n with the HAZ estimate of 0.09 standard deviations from column 1 of Table 6. The nutrition skill price P_n is calibrated from Hoddinott et al. (2011), who report that a one standard deviation increase in HAZ increases adult consumption by 20 percent. The effect of nutrition on test scores γ_{tn} is calibrated based on the observed correlation between test scores and nutrition in our sample (see Appendix Table A.10); this correlation seems likely to be an upward-biased estimate of the causal impact of nutrition on test scores, resulting in a conservative benefit-cost calculation.

The bottom panel of Appendix Table A.13 demonstrates that incorporating effects on HAZ boosts the benefit-cost ratio for the AWC sample substantially. When nutrition benefits based on HAZ scores are included, R increases from 12.2 to 17.5. In column 4 we instead measure nutrition based on stunting (HAZ below -2) and calibrate P_n with the Hoddinott et al. (2011) estimate that stunting reduces adult consumption by 66 percent. This results in a larger benefit-cost ratio of $R = 22.1$.

References

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly Journal of Economics* 126 (4), pp. 1593–1660.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch (2013). "Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education." *Education Policy in Developing Countries*, pp. 285–338.

Hoddinott, J., J. Maluccio, J. R. Behrman, R. Martorell, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, and K. M. Yount (2011). "The consequences of early childhood growth failure over the life course." (WP No. 01073). Washington, DC: IFPRI.

Kline, P. and C. R. Walters (2016). "Evaluating public programs with close substitutes: The case of Head Start." *The Quarterly Journal of Economics* 131 (4), pp. 1795–1848.