

NBER WORKING PAPER SERIES

AUGMENTING STATE CAPACITY FOR CHILD DEVELOPMENT:
EXPERIMENTAL EVIDENCE FROM INDIA

Alejandro J. Ganimian
Karthik Muralidharan
Christopher R. Walters

Working Paper 28780
<http://www.nber.org/papers/w28780>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2021

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Alejandro J. Ganimian, Karthik Muralidharan, and Christopher R. Walters. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Augmenting State Capacity for Child Development: Experimental Evidence from India
Alejandro J. Ganimian, Karthik Muralidharan, and Christopher R. Walters
NBER Working Paper No. 28780
May 2021
JEL No. C93,I21,I22,I25,O15

ABSTRACT

Despite growing interest in improving early-childhood education in developing countries, there is little evidence on cost-effective ways of doing so at scale. We use a large-scale randomized experiment to study the impact of adding an extra worker focused on pre-school education (for children aged 3-5) in the world's largest public early-childhood program: India's Integrated Child Development Services (ICDS). Adding a worker doubled net instructional time and led to 0.29 and 0.46 standard deviation increases in math and language test scores after 18 months for children who remained enrolled in the program. Rates of stunting and severe malnutrition were also lower in the treatment group, likely reflecting the effect of freeing up time of the incumbent worker to focus more on nutrition-related tasks. A cost-benefit analysis suggests that the benefits of the program are likely to significantly exceed its costs even under conservative assumptions.

Alejandro J. Ganimian
Steinhardt School of Culture, Education,
and Human Development
New York University
Kimball Hall 406W
246 Greene St.
New York, NY 10003-6677
alejandro.ganimian@nyu.edu

Christopher R. Walters
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
crwalters@econ.berkeley.edu

Karthik Muralidharan
Department of Economics, 0508
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER
kamurali@ucsd.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). Yet, despite growing policy interest in providing quality early-childhood education in developing countries around the world (World Bank, 2018), there is little evidence on cost-effective ways to improve early childhood education at scale.

This paper contributes evidence on this question 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 pre-school education. ICDS also caters to another 46 million children in the 0-3 age range with supplemental nutrition and health services provided through home visitation. ICDS services are provided through 1.35 million *anganwadi* centers (AWCs) across India by *anganwadi* workers (AWWs).¹ These services are provided free of charge and disproportionately cater to the poor. Thus, the *anganwadi* system 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 early-childhood education and nutrition programs.

Despite its importance, ICDS has limited staffing and funding, which may constrain its effectiveness. In particular, a single worker is responsible for health, nutrition, home visitation, and administrative tasks in addition to pre-school education, which limits time spent on pre-school education.² A preparatory scoping study we conducted in the state of Tamil Nadu (the setting of our study) found that workers spent only 38 minutes per day (out of a scheduled two hours) on pre-school instruction (Ganimian et al., 2018). 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 pre-school 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 *anganwadi* centers in four districts representative of a population of 60 million people in the Indian 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 focus on pre-school instruction,

¹We use the terms AWC and “center” interchangeably, and do the same with the terms AWW and “worker”.

²The worker is assisted by a helper who is responsible for cooking and cleaning. However, helpers are much less educated, and the worker is typically the only staff member with a secondary school education.

and was paid half the salary of a regular worker on a full-time shift. The intervention was implemented by the Government of Tamil Nadu using protocols similar to those that would be used if the program were to be scaled up, including hiring and training of the facilitators. The combination of random assignment in a representative sample of centers and government implementation of the program increases the external validity of our results to the full state of Tamil Nadu (as described in Muralidharan and Niehaus, 2017).

Our primary outcomes of interest are test scores in math, language, and executive function. We measure these through independent tests conducted in AWCs along with separate tests administered in a representative sample of households. We also measure child nutrition as a secondary outcome. Finally, we measure worker attendance, timeliness, and time on task through unannounced visits to the centers. Enumerators sat in classrooms for the full two hours scheduled for pre-school instruction (10am to 12pm) and coded all tasks performed by workers and facilitators in six-minute intervals. We present four sets of results on time use, education, nutrition, and cost-effectiveness respectively.

First, treated centers were closed only 3.8 percent of the time compared to the control group closure rate of 12.5 percent. On average, facilitators provided nearly an hour of daily pre-school instruction in treated centers during the two-hour observation window. Workers in treated centers reduced time spent on pre-school 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 pre-school education *doubled* in treated centers (from 38 to 76 minutes per day). Treated centers saw an over 250% increase 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.³

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

³Time spent on tasks other than education, health and nutrition, and administrative work is coded as “off task”. If a staff member is absent, the entire window is coded as “absent”. Thus the results on time use include both intensive margin effects and extensive margin effects (through changes in absence and closure rates). It is also possible that the total time spent on health and nutrition increased further outside our two-hour observation window (which was during the time scheduled for pre-school education), and that the facilitator helped with these tasks outside the observation window (see Section 4.2).

combination of children graduating out of pre-school, moving to private pre-schools, families migrating, and irregular attendance. We therefore supplement the AWC-based outcomes by also measuring outcomes for a representative sample of children from the baseline cohort in the household, 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 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 comparing household and center-based estimates for the common sample of children who took both tests. In this sample, we find household treatment effects of similar magnitude to those measured on the AWC tests. Moreover, treatment-on-the-treated effects obtained by scaling the household sample estimates by the share of students observed at the center are close to the AWC and common sample estimates. We therefore interpret the AWC estimates as reflecting treatment effects on children who actively attended *anganwadi* centers, while the household estimates capture intent-to-treat-style impacts on a broader population of 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 and boosted height-for-age z -scores (HAZ) by 0.09σ in AWC endline measurements. Children in treated centers were 3.1 percentage points less likely to be severely malnourished, defined as having a WAZ score below -3σ ($p < 0.01$). This represents a 34% reduction relative to the control mean of 9.1%. The treatment also reduced stunting, defined as having a HAZ score below -2σ , by 4.8 percentage points (a 16% reduction relative to the control mean of 29.1%, $p < 0.05$). 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 of the relationship between test-score effects of pre-school interventions and impacts on adult earnings, 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 12 times the cost. Incorporating the nutrition benefits for children in the AWC sample increases the estimated benefit/cost ratio to between 16 and 20. Since these calculations rely on strong assumptions, we conduct a sensitivity analysis, which suggests that the program would be cost-effective even under conservative assumptions regarding the economic value of test-score and nutrition gains. Moreover, our projections suggest that the government would fully recover program costs in present value if it captured as little as 6.3% 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.

Reflecting the growing academic consensus on the importance of investing in early-childhood development (Engle et al., 2007; Shonkoff, Radner, and Foote, 2017) policy documents around the world are increasingly highlighting the importance of providing quality early-childhood education, especially for low-income populations (e.g., GoI, 2020; World Bank, 2018). Yet, progress in translating these objectives into reality has been limited, in part because of the fiscal constraints of doing so. In particular, while evidence from smaller-scale studies of high-intensity programs suggests that investing in access and quality of early-childhood programs may make sense (Gertler et al., 2014; Heckman et al., 2010; Campbell et al., 2002), effects of such programs have been more modest at larger scales (Attanasio et al., 2014). As a result, policymakers may have reason to be cautious in funding large-scale expansions of early childhood education.

Thus, 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 (where governments can presumably achieve minimum standards of quality) have improved pre-primary school attendance and learning (Berlinski, Galiani, and Manacorda, 2008; Berlinski, Galiani, and Gertler, 2009). However, public preschool expansions in less-developed countries have been less effective, and successful programs in these settings have typically been operated by non-government entities (Martinez, Naudeau, and Pereira, 2013; Bouguen et al., 2014; Blimpo et al., 2019; Dean and Jayachandran, 2020). Our results suggest that strengthening existing public pre-school education systems by adding staff can be an effective option for improving early childhood education outcomes at scale.

Second, a large experimental literature has shown that interventions in the first 1,000 days of life (in-utero and up to age 2) can improve child nutrition (see, e.g., Britto et al., 2017). However, 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 between ages 3 and 5.⁴ 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 at scale.

⁴Barham, 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.

Third, and more generally, our results suggest that augmenting state capacity for service delivery in low-income countries by adding locally hired staff on modest salaries may be a highly cost-effective public investment. 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, 2017; de Ree et al., 2018). Thus, expanding hiring of local staff at lower than civil-service salaries may be a promising policy option for expanding state capacity for service delivery more broadly (Haines et al., 2007; Muralidharan, 2016).

Fourth, and related, 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 (DHHS, 2017; Berlinski and Schady, 2015; GoI, 2020). Our results suggest that such qualifications may not be necessary in settings with very low learning levels, since locally-hired staff with a secondary school education and just a week of training were highly effective at improving learning outcomes. These findings are consistent with similar results in the context of primary-school education in India and Kenya (Banerjee et al., 2007; Muralidharan and Sundararaman, 2013; Duflo, Dupas, and Kremer, 2015).⁵

Finally, we provide new evidence on the efficacy of increased resources in education production. A large literature has studied the impacts of smaller classes and other resource interventions on student outcomes, with mixed results (e.g., Krueger, 1999; Angrist and Lavy, 1999; Lazear, 2001; Fredriksson, Öckert, and Oosterbeek, 2013; Jackson, Johnson, and Persico, 2016; Lafortune, Rothstein, and Schanzenbach, 2018; Angrist et al., 2019; Jackson, Wigger, and Xiong, 2020). Our results demonstrate that in the ICDS context, increasing resources in the form of additional staff can generate large benefits. More generally, our estimates from India are consistent with historical evidence from higher-income countries that has found large benefits from investing in early childhood education and nutrition programs (Alex-Petersen, Lundborg, and Rooth, 2017; Hendren and Sprung-Keyser, 2020).

⁵However, this result may only apply to settings of very low student learning. Evidence from upper-middle income countries suggests that additional teacher qualifications (e.g., pre- or in-service training) may be needed in settings where most children have mastered basic skills (Andrew et al., 2019; Bernal et al., 2019). Even in our setting, more qualified pre-school teachers *may* deliver greater learning gains. However, our results suggest that it may not be necessary to insist on such qualifications to see meaningful positive effects. We discuss policy implications further in the Conclusion.

2 Setting and intervention

India has more young children than any other country in the world, with over 160 million children of age 0 to 6 (MHA, 2012). As per the latest internationally comparable data, 35% of Indian children are stunted and 33% are undernourished (UNESCO/WHO/World Bank, 2020).⁶ Data from a very recent national survey suggest that these figures may have worsened in recent years (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 (Grantham-McGregor et al., 2007; Lu, Black, and Richter, 2016).

Besides malnutrition, India also faces a severe challenge of low learning levels among children. For example, a recent survey representative of rural areas showed that 50% of students in fifth grade could not 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 (ASER, 2019). 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” (GoI, 2002). The Right to Education Act of 2009 promoted the free and public provision of education for children ages 3 to 6 (MLJ, 2009). The National Early Childhood Care and Education Policy Framework, adopted in 2013, stipulated that pre-school education should cover six broad domains of child development.⁷ 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” (GoI, 2020).

However, achieving these policy aspirations in practice has been difficult in part due to constraints in funding and state capacity for implementation. 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 policy level, but more so at the level of augmenting capacity for implementation.

⁶Children are considered stunted if their height-for-age is 2σ or more below World Health Organization (WHO) child-growth standards, and undernourished if their weight-for-age is 2σ or more below WHO standards. They are considered severely stunted or undernourished if these figures are 3σ or more below WHO standards.

⁷These include sensory and perceptual development; physical and motor development; language development; cognitive development; creative and aesthetic appreciation; and personal, social, and emotional development (MWCD, 2013).

2.1 The Integrated Child Development Services

The Integrated Child Development Services (ICDS) is the main public program through which the Government of India promotes early-childhood development in India. ICDS is the largest single provider of early childhood health and education services in the world (UNESCO, 2019). It caters to over 36 million 3- to 6-year-olds, and aims to provide a range of early-childhood health and nutrition services as well as pre-school education.⁸ It also caters to another 46 million children in the 0-3 age range, through home visitation services and provision of supplemental nutrition packets.

ICDS provides all of its services through *anganwadi* centers (AWCs). Each *anganwadi* center serves a catchment area of 400-800 people (PEO, 2011), and is typically staffed with one *anganwadi* worker (AWW) and one *anganwadi* helper. The worker is responsible for managing all services provided at the center, spanning early childhood health, nutrition, pre-school 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; monitoring children’s nutritional status and 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.⁹ 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).¹⁰ 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. *Anganwadi* workers 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 secondary school (10th grade), which is a meaningful marker of human capital since it requires passing an external exam (ICDS, 2017).

The helpers serve as assistants to the workers. They are responsible for 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, cooking, and distributing meals

⁸For comparison, the entire Head Start program in the US (which has been studied much more intensively) had 652,422 funded slots as of 2019, which is less than 2% of the coverage of the ICDS (NHSA, 2020).

⁹The full list of responsibilities of an *anganwadi* worker is available at: <https://bit.ly/2YV73iB>. Last accessed: February 3, 2021.

¹⁰These registers include: family details, supplementary food stock, supplementary food distribution, pre-school education, pregnancy and delivery, home visits planner, referrals, and office stock (daily); immunization and village health and nutrition day, vitamin A, weight records of children, severe and acute malnutrition, and tracking of state-specific programs for pregnant and lactating women (monthly). There is an ongoing effort by both the Government of India and state governments to digitize data collection through smartphones, but the default remains paper-based record keeping by *anganwadi* workers.

and nutritional supplements at the center and to eligible households in the catchment area (under the supervision of the worker). 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).

2.2 The early-childhood education (ECE) facilitator intervention

Our study is set in the southern Indian state of Tamil Nadu. Tamil Nadu is India’s sixth most populous state and one of the more developed states in the country, scoring in the top third of a national index of human development (IAMR, 2011). It is home to 4.2 million children aged 3 to 6, or 4% of children in this age group nationwide (MHA, 2012). Child-nutrition outcomes are better than national averages, but still concerning by absolute standards: 27% of children are stunted and 24% are undernourished (MHFW, 2017).

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 pre-school education as a promising idea to consider. We also conducted a diagnostic study on worker time-use in 24 centers and found that workers spent only 38 minutes per day on pre-school instruction on average (Ganimian et al., 2018). Further, in surveys conducted for the 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.¹¹

The intervention we study provided randomly-selected centers the opportunity to hire an extra early-childhood education (ECE) facilitator to focus on pre-school 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 pre-school education from 10am to 12pm. Facilitators were expected to work half the hours of an AWW, and were correspondingly paid around half the salary (Rs. 4000/month compared to Rs. 8000/month).

GoTN had already developed instructional content for early-childhood education as well as protocols for training workers in implementing this curriculum. This curriculum was designed

¹¹Tamil Nadu had in fact implemented a 2-worker model in the ICDS in the 1980s under the Tamil Nadu Integrated Nutrition Project funded by the World Bank and other donors (see Heaver, 2002). Qualitative evaluations and interviews suggested that the program was well received at the field level and effective. However, the second worker was discontinued when donor funding for the project ended. This institutional memory of having implemented a 2-worker model in the past also contributed to the government’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.

in partnership with UNICEF, and reflected global guidelines for early-childhood education expressed in locally contextualized content. GoTN developed manuals for facilitators based on its materials for AWWs, and provided them with one week of training on implementing this curriculum. It also trained facilitators and workers on their expected division of labor before the start of each school year. GoTN’s communications to field staff noted that the goal of the program was to both improve the quantity and quality of educational 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.¹² Thus, the intervention did not change any of the stated goals of the *anganwadi* centers, but augmented their capacity to deliver these goals.

The addition of the facilitator could improve outcomes in several ways, including (a) increasing the likelihood of centers opening on time and reducing the chances of their being closed; (b) increasing pre-school instructional time during working hours; (c) enabling instruction in smaller groups if the worker and facilitator teach simultaneously (equivalent to a class-size reduction); and (d) increasing workers’ time available for health and nutrition related tasks. Our results should therefore be interpreted as the composite effect of the intervention through these various channels.

3 Research methods

Our design and methods follow a registered pre-analysis plan.¹³

3.1 Sampling

We randomly sampled four districts across the state, to be representative of a population of 60 million people.¹⁴ 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 each (PEO, 2011). We started with the universe of centers in the four sampled districts and excluded centers with other NGO interventions, those in buildings shared with other centers, and those with vacancies in both staff positions (worker and helper).¹⁵ We then randomly sampled 320 centers from the remaining universe, stratifying by staffing vacancy, and project.

¹²All control and treatment centers were also provided data on nutritional status of children in the center based on the independent baseline measurements carried out by the research team.

¹³See <https://www.socialscienceregistry.org/trials/1772>.

¹⁴Tamil Nadu has a population of 67 million. We excluded the district of Chennai, which is the state capital and a metropolis of 7 million people. The sample is thus representative of the state, excluding Chennai. District sampling was stratified by geographic zones and average nutrition status. The four sampled districts were Kancheepuram, Karur, Trichy, and Virudhunagar.

¹⁵Together, this resulted in dropping 10.8% of centers from the sampling frame.

3.2 Randomization and summary statistics

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.¹⁶ 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.¹⁷

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. Children attending the centers were 3.5 years old on average. Baseline nutrition levels were low: 37% of children were severely or moderately underweight with a weight-for-age z -score (WAZ-score) below -2, and 35% were severely or moderately stunted with a height-for-age z -score (HAZ-score) below -2. Consistent with *anganwadi* centers enrolling children from relatively disadvantaged families, the fraction of underweight children in our sample is higher than the state-wide rate of 26% in rural Tamil Nadu (MHFW, 2017).

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 student 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.

3.3 Implementation quality

GoTN implemented the intervention well. Virtually all treatment centers hired an early childhood education facilitator. In our first process monitoring survey, conducted five months after GoTN issued the notification to treated centers to hire the facilitator, 98% of centers had a facilitator (Table A.1). On average, facilitators were 29 years old and had been hired 135 days prior to the survey, confirming that they were hired promptly, within 15-30 days of

¹⁶These included population, age distribution, language, occupation distribution, and family income based on administrative data for each AWC catchment area.

¹⁷This 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 all focused on improving child nutrition and results are reported in a companion paper (Muralidharan, Walters, and Ganimian, 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).

the notification being issued. Further, nearly all facilitators (96%) reported having received the initial 6-day training required by the government.

3.4 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.¹⁸ 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 B 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 quite sensitive to field protocols, enumerators received extensive training, and each child was measured twice. Appendix B 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 being insignificant (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.

However, the overall follow-up rate in the AWC sample (of children in the baseline) was only 33%. This likely reflects a combination of children moving out of the *anganwadi* centers to private pre-schools, ageing out of pre-school and enrolling in first grade at age 5, migrating, and attending irregularly (we see that older children are significantly less likely to be in the AWC follow-up sample). Data from a different study in the same districts show the age profile of student attendance, and we see that the proportion of students enrolled in AWCs drops substantially from age 3 to age 5 (Figure A.1).¹⁹ Since there was no differential attrition, we

¹⁸We 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.

¹⁹The data used for Figure A.1 is from a different study in the same districts (but from different villages) that aims to study household choice behavior across pre-school options, and therefore collected much more

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 students who were observed in the baseline sample, visited their households, and tested them there in May-June 2018.²⁰ There was again no differential attrition between treatment and control groups overall, or by observable baseline characteristics (Table 2, col. 2). However, the followup 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 pre-school 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. The announced visits were used to survey workers and facilitators and to obtain additional details on teaching practices.

3.5 Estimation

We estimate program impacts by comparing mean outcomes for children in treatment and control centers, controlling for randomization strata and baseline characteristics. The main estimating equation is:

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

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 students, the mean baseline outcome for all students 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. The parameter of interest is β , which represents the average causal effect of the a center receiving the early-childhood education facilitator intervention.

detailed household survey data on the *universe* of children in these villages. We thank Mauricio Romero and Abhijeet Singh for sharing the data for this figure.

²⁰To ensure sufficient precision for studying effects on malnourished children, we oversampled students with WAZ scores below -2σ for the household survey. All results that use this sample are re-weighted to be representative of the full baseline sample.

We estimate equation (1) by ordinary least squares (OLS) regression in the AWC sample. Regressions for the household sample are weighted to account for differences in sampling probabilities for the household survey, allowing us to recover effects for the population of children who took the baseline test. Standard errors are clustered at the center level.

4 Results

4.1 Center openings and staff attendance

The addition of the early-childhood education facilitator significantly reduced the likelihood of centers being closed at the scheduled start of pre-school 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).²¹ 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 students), centers are typically locked with the AWW 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.²²

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 pre-school instruction (10am-12pm). We find that facilitators spent around half this time (57 minutes) on pre-school instruction (Table 4, col. 1). They spent around 20 minutes on administrative work, and 6

²¹The 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.

²²This finding contrasts with that of Muralidharan and Sundararaman (2013) and Duflo et al. (2015) who find that adding a contract teacher to schools in India and Kenya reduced attendance of the existing teachers.

minutes on health and nutrition tasks. The remaining 37 minutes were either off-task (27 minutes) or accounted for by absence (10 minutes).²³

The intervention also shifted the time allocation of *angwanwadi* workers in the expected direction. Worker time spent on pre-school 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 treatment 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 vs. 22 minutes), and on health and nutrition tasks (11 vs. 6 minutes).²⁴ 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.²⁵

Despite the reduction in instruction 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 pre-school 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).²⁶ 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 pre-school education, health and nutrition, and administrative work (col. 6).

Since the time window we observe was the part of the day scheduled for pre-school instruction, it is unsurprising that the biggest absolute impact on time spent was on pre-school 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

²³When 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.

²⁴The health and nutrition category captures time spent preparing or serving food, assisting children to use the toilet, and miscellaneous health-related activities (see Table A.2 for details).

²⁵It 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 the interaction provided parents with inputs and advice on home interactions with children. We are not able to quantify this since enumerators did not code the identity of who the staff were interacting with.

²⁶Tables A.3-A.6 display impacts on additional measures of time allocation within each domain. Based on announced observations, we see that roughly half of the extra pre-school instructional time was spent on “practice and drill” activities, with the rest split between classwork and question and answer time (Table A.6).

suggests that they spent around 1.5 hours per week on health-related activities (Table A.5), substantially more than the six minutes per day we see in the observation window. Second, the 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 pre-school instruction, the provision of an early childhood education facilitator produced large test-score gains. Students in treated centers 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, we find that students in treated centers scored 0.29σ higher ($p < 0.01$).²⁷

Treatment effects on the household tests were smaller but still significant. As shown in the second row of Table 5, students in treated centers scored 0.13σ higher in math ($p < 0.01$), 0.10σ higher in language ($p < 0.05$), and 0.05σ higher in executive function (not significant) on household assessments. On the composite measure of learning, students in treated centers scored 0.11σ higher ($p < 0.05$).

The difference in treatment effects between the AWC and household assessments most likely reflects the fact that the household sample includes students who were in the AWC at baseline, but were not attending at the endline. Figure A.2 examines this possibility by plotting the age distribution of children in the baseline and in both the AWC and household follow-up samples. Figure A.2 clearly shows that the household sample includes many more children over 5, who are likely to have aged out of the AWC (as shown in Figure A.1). Thus, the lower estimated effects in the household sample likely reflects 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 we see no meaningful

²⁷Table A.8 shows that these results are similar for outcomes defined as the proportion of test items answered correctly rather than standardized scores.

difference in treatment effects on the AWC and household tests in the common sample (0.32σ vs. 0.29σ for composite scores).²⁸

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, and the smaller effects in the household sample as reflecting an intent-to-treat (ITT) estimate on students present in the AWC during the baseline. If we assume that all treatment effects in the household sample accrue to students 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 ($0.29-0.31\sigma$). 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 students in the household sample who were not captured in the AWC assessments may have received some benefits from the intervention.²⁹

Finally, we find limited evidence of heterogeneity in treatment effects across achievement levels or baseline child characteristics. We plot treatment and control composite score distributions along with quantile treatment effects for the AWC assessments, and see that the treatment distribution first-order stochastically dominates the control distribution, suggesting broad-based test-score gains from the program (Figure 1, Panel A). Panel B plots non-parametric estimates of average treatment effects at each percentile of baseline composite score, and shows that there were large positive impacts across the full range of baseline achievement. We also find no evidence of differential effects by baseline nutrition (weight-for-age z -score), mothers' education, or AWW vacancy (Table A.11). 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

Our pre-analysis plan stated that nutrition is a secondary outcome for this study, since the intervention primarily targeted improving pre-school education through adding an early childhood education facilitator, and any impacts on nutrition would be an indirect consequence of freeing up AWW time to focus more on nutrition. However, consistent with the data on time use showing a near tripling of total time spent on health and nutrition activities (albeit from a

²⁸As described in Appendix B, the household assessment was shorter and included a subset of items from the AWC assessment. Table A.9 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.

²⁹Similarly, Table A.10 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.

modest base of six minutes per day),³⁰ and evidence that centers were significantly less likely to be closed, we find significant improvements in nutrition based on endline measurements at the AWC on both WAZ and HAZ scores.

Table 6 shows that average WAZ scores were 0.10σ higher in the treatment group ($p < 0.01$). The nutrition literature has noted that the mapping from WAZ score to long-term health outcomes may be highly non-linear, with improvements mattering much more for undernourished children (Hoddinott et al., 2013). We therefore also report effects on the incidence of moderate and severe malnutrition (WAZ scores below -2 and -3). We find no significant reduction in the probability that children are moderately underweight, but find a significant 3.1 percentage point decline in the rate of severe malnutrition ($p < 0.05$). This represents a 34% reduction in severe malnutrition relative to the control mean of 9.1%.

As with the test-score results, the impacts on nutrition 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 of children in both samples (Panel B), we find broadly similar increases in average WAZ in the AWC and household measurements (0.12σ and 0.07σ). Results for severe malnutrition are directionally similar, though the household estimate is not statistically significant. Finally, we investigate distributional treatment effects on WAZ scores in the AWC measurements and find broad-based evidence of positive impacts, with little systematic evidence of effect heterogeneity (see Figure 2 and Table A.12).

Similar patterns are observed for height-for-age z -scores. Average HAZ scores in treated centers exceed those in control AWCs by 0.09σ (Table 7, Panel A). The treatment group also saw a 16% reduction in the fraction of children who were stunted (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 in the case of the WAZ results, we see positive but insignificant estimates for HAZ outcomes in the household sample, and find generally comparable effects on AWC and household measurements when restricted to the common sample of children observed in both settings (Panel B). Figure 3 suggests increases in HAZ scores throughout the distribution, while Table A.13 shows similar impacts across subgroups. We also verify that the nutrition results are not sensitive to measurement outliers. Tables A.14-A.17 show that the estimated WAZ and HAZ impacts are robust to dropping outlier measurements and winsorizing the outcome variables.

The positive impacts on nutrition in the AWC sample reflect the experience of children who regularly attended the centers and had not attrited from the sample over the 18 months

³⁰As noted in Section 4.2, total staff time spent on health and nutrition tasks could have also increased further outside the observation window.

of the study.³¹ These results suggest that the benefits of providing an extra worker to focus on educational activities was not restricted to improving education outcomes, but extended to improving nutrition outcomes as well. 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 attend regularly.

Finally, we consider the extent to which improved nutrition outcomes may have contributed to the improvement in learning outcomes we find. In the baseline data, the coefficient from a regression of composite test scores on WAZ scores equals 0.15, while the corresponding coefficient for HAZ scores equals 0.14 (see Table A.20). These observational associations 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 early childhood education 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 being a second-order channel for learning gains during the study period.

5 Cost-effectiveness

We now assess how the estimated benefits of adding the early childhood education facilitator compare against the cost of doing so. 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 to project the impacts of the intervention on future earnings. Specifically, we use estimates collected in Kline and Walters (2016) for the long-term labor market returns to improving early-childhood learning and estimates from Hoddinott et al. (2013) for the long-term labor market returns to reducing early childhood malnutrition and stunting. Comparing the present discounted value of the projected increases in participant earnings to the intervention’s cost yields a simple economic benefit/cost metric for the program. Since this exercise relies on strong assumptions to forecast impacts into adulthood, we also conduct a sensitivity analysis that jointly varies each of the parameters we calibrate over a wide range of possible values.

The results of this cost effectiveness analysis are presented in Table 8. Panel A lists the assumptions used to calculate the expected present discounted value of future earnings for

³¹Similar to the test-score results, Tables A.18 and A.19 show more positive WAZ and HAZ estimates in the household sample for those who self-report AWC attendance.

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 Social Survey (NSS). We assume that workers are employed for 250 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 (Ministry of Statistics and Programme Implementation, 2021). We assume a working age from age 22 to age 65 and discount these projections back to the present using a discount rate of 3 percent.³² These parameter values imply that the average present discounted value (PDV) of total future earnings for children in rural Tamil Nadu equals roughly INR 3.3 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, 2020; 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.

Column 1 of Table 8 uses the full household survey sample and predicts earnings gains using only test-score impacts, ignoring any effects on nutrition since the nutrition estimates for this sample are insignificant. Assuming 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 47,000 per child.

Column 1 of Panel C uses this projection to compute a benefit/cost ratio for the early childhood education facilitator intervention, calculated as the ratio of earnings gains to program costs. Our experimental sample included 14 children per treated center at baseline. Since the experiment was conducted over eighteen 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

³²The 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 is 2-3%. Current yields on 10-year government bonds are around 6-7%, whereas the inflation rate is 4-5%.

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 by 14×1.33 to calculate a total expected gain from the program, which equals about INR 882,000. The cost of the program was roughly INR 74,000 per center over eighteen months.³³ Taking the ratio of these two numbers produces a predicted benefit/cost ratio of 11.8.

The remaining columns of Table 8 repeat this calculation using estimates from the AWC sample rather than the household sample to predict the long-run impacts of the intervention. In this sample, the test-score effect is larger (0.29σ vs 0.11σ), but the number of children who see the gains is smaller (5 vs. 14), reflecting the smaller number of children who remained in the AWC sample. Put together, using estimates from the AWC sample yields a benefit/cost ratio of 11.1 (column 2), which is very similar to the 11.8 estimated in the household sample.

Incorporating nutrition gains (which are significant for the AWC sample) boosts the benefit-cost ratio substantially (columns 3-4). Hoddinott et al. (2011) study the long-term links between improving childhood nutrition and adult labor-market outcomes in Guatemala, and estimate that a 1σ improvement in childhood HAZ score generates a 20% increase in household consumption per capita in adulthood. Following Hoddinott et al. (2013), we use this estimate as a proxy for impacts on adult income. Applying this number to the estimated impact of the intervention on HAZ scores (0.09σ ; Table 7, col. 1) and adjusting for the fact that some of the nutrition impact likely operates through test scores,³⁴ we project that the combined earnings impact of the intervention through both test scores and nutrition equals about INR 179,000, which is 43 percent larger than the projection based on test scores alone. Column 3 shows that incorporating nutrition boosts the benefit/cost ratio from 11.1 to 16.0. Since the benefits of nutrition may be highly nonlinear, column 4 also reports a version of this calculation that attributes all nutrition gains to reductions in stunting, which results in a somewhat larger benefit/cost ratio (20.2 rather than 16.0).

Our cost/benefit analysis relies on strong assumptions and estimates from other contexts to project the economic value of test score and nutrition gains. We therefore test the sensitivity of our results to these assumptions by varying each of the calibrated parameters in Table 8. Specifically, we draw each unknown parameter from a distribution defined on a range of possible values with the values from Table 8 in the center. Figure 4 presents the results of this sensitivity analysis. Panel A presents distributions of benefit/cost ratios corresponding to the scenarios in column 1 of Table 8 (using only test-score gains in the household sample),

³³This cost includes a one-time training cost for each facilitator along with 18 months of salary and administrative costs for the program.

³⁴We make this adjustment by subtracting the predicted impact of the nutrition increase on test scores from the total test score impact before applying the earnings gain per unit to each of the two treatment effects. This prediction is based on the slope of 0.14 test score standard deviations per unit of HAZ observed in our sample (Table A.20, Panel B, col. 4).

and Panel B does so based on column 3 (using test-score and nutrition gains in the AWC sample). In each panel the left-hand plot draws parameters independently from a uniform distribution and the right-hand panel draws them from a truncated normal distribution. 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 4-34. For Panel B, the range is 6-45. Thus, even under very conservative realizations of the parameters used to value test score and nutrition gains, the benefits are likely to be greater than the costs.³⁵

Further, the calculations above are also conservative in not including the non-labor market benefits of better health and education. These include the intrinsic value of better health and education as a direct determinant of 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., Heckman, Stixrud, and Urzua, 2006; Vogl, 2012; Cole, Paulson, and Shastry, 2016).

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 8 indicate that as long as children pay a net tax rate of at least $(74,478/1,190,346) \times 100 = 6.3\%$ on future earnings, program costs are more than paid back in present value by expected future tax revenues.³⁶ In a historical analysis of returns to social programs in the United States, Hendren and Sprung-Keyser (2020) report that early-childhood interventions are among the most cost-effective and frequently generate earnings gains sufficient to recover the government's costs. Our results suggest that such returns may be possible for contemporary large-scale early childhood interventions in India.

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, 2018). Yet, despite broad agreement on its importance among both academic experts and in global and national policy documents, there is remarkably little evidence on how

³⁵Thus, even if states were to hire a second AWW at the regular pay of an AWW (which is double the pay 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.

³⁶Even 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 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 6.3% of increases in expenditure as tax revenue.

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 no 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 pre-school 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 pre-school 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 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. We estimate that the present discounted value of the projected increase in earnings from improving early-childhood test scores is 11-12 times greater than the cost. Adding the present discounted value of the increased earnings from the nutrition gains delivered by the program increases the benefit-cost ratio to between 16 and 20.

Beyond the program benefits and cost effectiveness, the case for scaling up this approach across the ICDS system is strengthened by its ease of implementation, and the speed with which it can be scaled up. The treated centers were able to successfully hire an extra early childhood education facilitator within 15-30 days of being allowed to do so, with no additional support beyond the financial authorization to do so. More generally, the expansion of school education across India over the last two decades has also expanded the pool of secondary-school educated young women who form the pool of eligible candidates for the role. Thus, the supply of candidates is unlikely to be a constraint to scaling up.

Further, constraints on traveling outside their village is an important barrier for labor-force participation for young women in rural South Asia (Andrabi, Das, and Khwaja, 2013). Thus, expanding the approach studied in this paper, which provides employment in the same village, 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). Thus, scaling up the program we study may have additional benefits for women’s empowerment and education beyond the direct benefits on early childhood education and nutrition.

One limitation of our analysis is that our estimates of cost effectiveness rely on links between short-term and long-term impacts estimated from other settings (Hoddinott et al., 2011; Kline and Walters, 2016). While recent evidence confirms that successful early health

and education interventions in developing countries can have persistent long-term positive impacts (Miguel and Kremer, 2004; Baird et al., 2016), we cannot be sure that these mappings are stable across interventions and contexts.

The paucity of evidence linking short- and long-term impacts may reflect weak incentives for policymakers, funding bodies, and researchers to conduct long-term studies that require extended waiting periods before producing results (Budish, Roin, and Williams, 2015). This highlights the importance of longitudinal studies linking treatment effects on early-life outcomes to outcomes at various later points in life for children in developing countries (e.g., Attanasio, Meghir, and Nix, 2020), as well as the importance of developing and validating methods for using short- and medium-term outcomes to construct accurate forecasts of long-term impacts (Athey et al., 2019). In future work, we hope to contribute evidence on these questions by studying the impacts of the early childhood education facilitator intervention on longer-run educational, health, and economic outcomes.

References

- Alex-Petersen, J., P. Lundborg, and D.-O. Rooth (2017). “Long-term effects of childhood nutrition: Evidence from a school lunch reform.” (IZA DP No. 11234). Bonn, Germany: Institute of Labor Economics (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 Working Paper No. 26191). Cambridge, MA: NBER.
- Angrist, Joshua D and Victor Lavy (1999). “Using Maimonides’ rule to estimate the effect of class size on scholastic achievement.” *The Quarterly journal of economics* 114 (2), pp. 533–575.
- Angrist, Joshua D, Victor Lavy, Jetson Leder-Luis, and Adi Shany (2019). “Maimonides’ Rule Redux.” *American Economic Review: Insights* 1 (3), pp. 309–24.
- ASER (2019). “Annual status of education report (rural) 2018.” New Delhi, India: ASER Centre.
- (2020). “Annual status of education report (rural) 2019: Early years.” New Delhi, India: ASER Centre.
- Athey, Susan, Raj Chetty, Guido W Imbens, and Hyunseung Kang (2019). *The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely*. Tech. rep. National Bureau of Economic Research.
- Attanasio, O. P., C. Fernández, E. O. A. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014). “Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: Cluster randomized controlled trial.” *BMJ* 349, g5785.
- Attanasio, O. P., Costas Meghir, and Emily Nix (2020). “Human capital development and parental investment in India.” *The Review of Economic Studies* 87 (6), pp. 2511–2541.
- Bailey, M. J., B. D. Timpe, and S. Sun (2020). “Prep school for poor kids: The long-run impacts of Head Start on human capital and economic self-sufficiency.” (NBER Working Paper No. 28268). Cambridge, MA: NBER.
- Baird, S., J. H. Hicks, M. Kremer, and E. Miguel (2016). “Worms at work: Long-run impacts of a child health investment.” *The Quarterly Journal of Economics* 131 (4), pp. 1637–1680.
- Banerjee, Abhijit V., Shawn 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 (2017). “The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers.” *Unpublished manuscript*. Washington, DC: The World Bank.
- 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.

- 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 (IDB).
- Bernal, R., O. P. Attanasio, X. Peña, and M. Vera-Hernández (2019). “The effects of the transition from home-based childcare to childcare centers on children’s health and development in Colombia.” *Early Childhood Research Quarterly* 47 (2), pp. 418–431.
- Blimpo, M. P., P. Carneiro, P. Jervis, and T. Pugatch (2019). “Improving access and quality in early childhood development programs: Experimental evidence from The Gambia.” (Policy Research Working Paper No. 8737). Washington, DC: The World Bank.
- Bouguen, A., D. Filmer, K. Macours, and S. Nadeau (2014). “Preschool and parental response in a second best world: Evidence from a school construction experiment.” *The Journal of Human Resources* 53 (2), pp. 474–512.
- Britto, Pia R, Stephen J Lye, Kerrie Proulx, Aisha K Yousafzai, Stephen G Matthews, Tyler Vaivada, Rafael Perez-Escamilla, Nirmala Rao, Patrick Ip, Lia CH Fernald, et al. (2017). “Nurturing care: promoting early childhood development.” *The Lancet* 389 (10064), pp. 91–102.
- Budish, Eric, Benjamin N Roin, and Heidi Williams (2015). “Do firms underinvest in long-term research? Evidence from cancer clinical trials.” *American Economic Review* 105 (7), pp. 2044–85.
- Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002). “Early child-hood education: Young adult outcomes from the Abecedarian Project.” *Applied Developmental Science* 6 (1), pp. 42–57.
- Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny 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, Raj, John N Friedman, and Jonah 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, Shawn, Anna Paulson, and Gauri Kartini 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.
- 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, Joshua T and Seema Jayachandran (2020). *Attending kindergarten improves cognitive but not socioemotional development in India*. Tech. rep. Working Paper.
- Department of Health and Human Services, Administration for Children and Families (2017). “Head Start program performance standards. 45 CFR chapter XIII. Preamble - Part I.” Washington, DC: Department of Health and Human Services, Administration for Children and Families.
- Duflo, E., J. Berry, S. Mukerji, and M. Shotland (2015). “A wide angle view of learning: Evaluation of the CCE and LEP programmes in Haryana, India.” (Impact Evaluation Report No. 22). International Initiative for Impact Evaluation (3ie). New Delhi, India.

- Duflo, E., P. Dupas, and M. Kremer (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)*, pp. 467–514. Oxford, UK: North Holland.
- Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek (2013). “Long-term effects of class size.” *The Quarterly Journal of Economics* 128 (1), pp. 249–285.
- Ganimian, A. J., A. Venkatachalam, S. Gade, and K. Muralidharan (2018). “Too little, too early: Instructional time during pre-school education in India.” New Delhi, India: J-PAL.
- Garces, Eliana, Duncan Thomas, and Janet 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.
- Gertler, P. J., J. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. M. Grantham-McGregor (2014). “Labor market returns to an early childhood stimulation intervention in Jamaica.” *Science* 344 (6187), pp. 998–1001.
- Government of India (GoI) (2002). “The Constitution (Eighty-sixth Amendment) Act.” URL: <http://indiacode.nic.in/coiweb/amend/amend86.htm> (Last accessed: April 26, 2017). Government of India. New Delhi, Delhi.
- (2020). “National Education Policy 2020.” New Delhi, India: Ministry of Human Resource Development, Government of India.
- Government of Tamil Nadu (GoTN) (2021). “Criteria for the appointment of *anganwadi* worker/mini *anganwadi* worker/helper.” URL: https://icds.tn.nic.in/appointment_criteria.html. Last accessed: January 20, 2021.
- Grantham-McGregor, S. M., Y. B. Cheung, S. Cueto, P. Glewwe, L. M. Richter, B. Strupp, International Child Development Steering Group, et al. (2007). “Developmental potential in the first 5 years for children in developing countries.” *The Lancet* 369 (9555), pp. 60–70.
- 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.
- Heaver, R. (2002). “India’s Tamil Nadu Nutrition Program: Lessons and Issues in Management and Capacity Development.” Washington, DC: The 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., S. H. Moon, P. A. Pinto R. Savelyev, and A. Yavitz (2010). “The rate of return to the Highscope Perry Preschool Program.” *Journal of Public Economics* 94 (1-2), pp. 114–128.

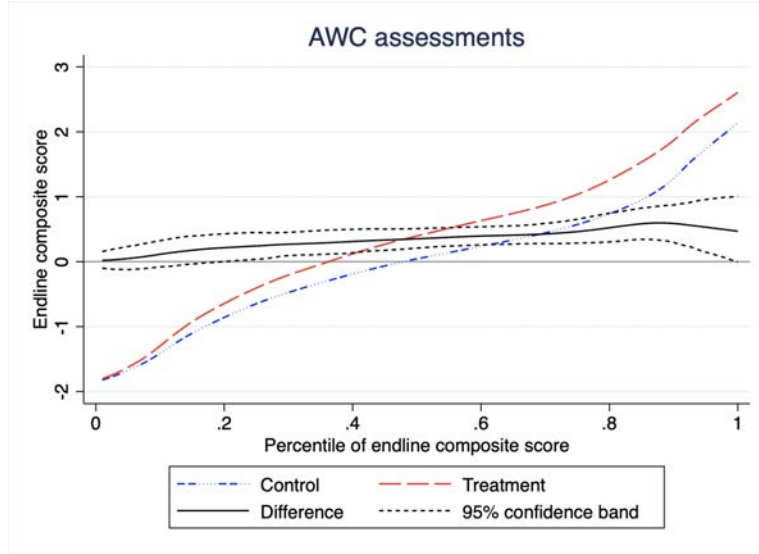
- Heckman, J. J. and S. Mosso (2014). “The economics of human development and social mobility.” *Annual Review of Economics* 6, pp. 689–733.
- Heckman, J. J., J. Stixrud, and S. Urzua (2006). “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.” *Journal of Labor Economics* 24 (3), pp. 411–482.
- 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.” (IFPRI Discussion Paper No. 01073). Washington, DC: International Food Policy Research Institute (IFPRI).
- Institute of Applied Manpower Research (IAMR) (2011). “India Human Development Report 2011: Towards Social Inclusion.” URL: http://www.im4change.org/docs/340IHDR_Summary.pdf (Last accessed: May 1, 2017). Planning Commission. Government of India. New Delhi.
- Integrated Child Development Services (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.
- Jackson, C. K., R. C. Johnson, and C. Persico (2016). “The effects of school spending on educational and economic outcomes: Evidence from school finance reforms.” *The Quarterly Journal of Economics* 131 (1), pp. 157–218.
- Jackson, C. K., C. Wigger, and H. Xiong (2020). “Do school spending cuts matter? Evidence from the Great Recession.” *American Economic Journal: Economic Policy*.
- Jensen, Robert (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.
- Krueger, A. B. (1999). “Experimental estimates of education production functions.” *The Quarterly Journal of Economics* 114 (2), pp. 497–532.
- Lafortune, J., J. Rothstein, and D. W. Schanzenbach (2018). “School finance reform and the distribution of student achievement.” *American Economic Journal: Applied Economics* 10 (2), pp. 1–26.
- Lazear, E. P. (2001). “Educational production.” *The Quarterly Journal of Economics* 116 (3), pp. 777–803.
- 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 (2013). “Preschool and child development under extreme poverty: Evidence from a randomized experiment in rural Mozambique.” (Policy Research Working Paper No. 8290). Washington, DC: The World Bank.

- Miguel, E. and M. Kremer (2004). “Worms: identifying impacts on education and health in the presence of treatment externalities.” *Econometrica* 72 (1), pp. 159–217.
- Ministry of Health and Family Welfare (MHFW) (2017). “National family health survey-4 (2015-16). State fact sheet: Tamil Nadu.” New Delhi, India: Government of India.
- (2020). “National family health survey-5 (2019-20). Key indicators from 22 states/UTs from phase I.” New Delhi, India: Government of India.
- Ministry of Home Affairs (MHA) (2012). “15th Census of India.” URL: <http://www.censusindia.gov.in/> (Last accessed: February 15, 2016). Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, Government of India. New Delhi, Delhi.
- Ministry of Law and Justice (MLJ) (2009). “The right of children to free and compulsory education act, 2009.” New Delhi, India: Government of India.
- Ministry of Statistics and Programme Implementation (2021). “Indian states by GDP per capita growth.” URL: <https://statisticstimes.com/economy/india/indian-states-gdp-per-capita-growth.php>. Last accessed: February 27, 2021.
- Ministry of Women and Child Development (MWCD) (2013). “National early childhood care and education policy.” New 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., M. Romero, and K. Wüthrich (2021). “Factorial designs, model selection, and (incorrect) inference in randomized experiments.” (NBER Working Paper No. 26562). Cambridge, MA: NBER.
- Muralidharan, K. and V. Sundararaman (2013). “Contract teachers: Experimental evidence from India.” (NBER Working Paper No. 19440). Cambridge, MA: NBER.
- Muralidharan, K., C. R. Walters, and A. J. Ganimian (2020). “Improving school preparedness and child health outcomes through Integrated Child Development Services (ICDS) in Tamil Nadu.” *Unpublished manuscript*. New Delhi, India: J-PAL.
- NHSA (2020). “2020 national Head Start profile.” Alexandria, VA: National Head Start Association (NHSA).
- Pritchett, L. (2009). “Is India a flailing state? Detours on the four lane highway to modernization.” (HKS Faculty Research Working Paper Series No. RWP09-013). Cambridge, MA: Harvard Kennedy School of Government.
- Program Evaluation Organization (PEO) (2011). “Evaluation report on Integrated Child Development Services.” New Delhi, India: Planning Commission, Government of India.
- 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.
- Shonkoff, J. P., J. M. Radner, and N. Foote (2017). “Expanding the evidence base to drive more productive early childhood investment.” *The Lancet* 389 (10064), pp. 14–16.
- Singh, A. (2014). “Giving up too early on malnourished children? Catch-up growth and Midday Meals.” *Ideas for India*. URL: <https://www.ideasforindia.in/topics/macroeconomics/giving-up-too-early-on-malnourished-children-catch-up-growth-and-midday-meals.html>. Last accessed: April 26, 2021.

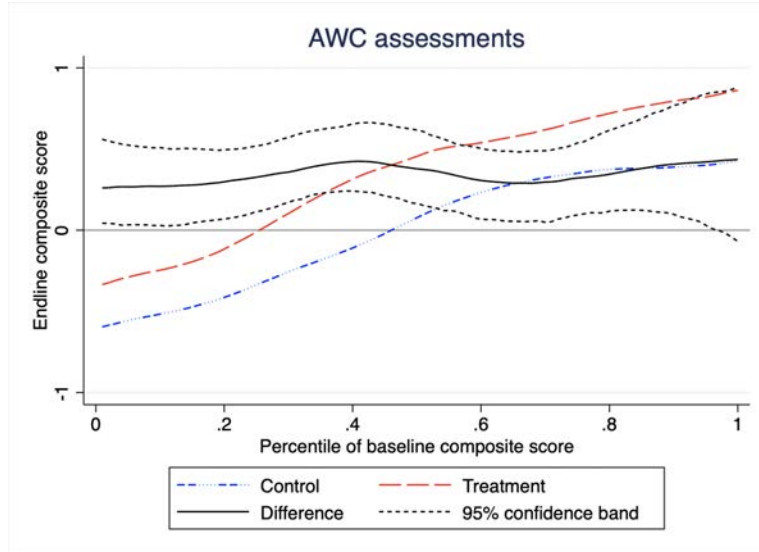
- Singh, A., A. Park, and S. Dercon (2014). "School meals as a safety net: An evaluation of the Midday Meal scheme in India." *Economic Development and Cultural Change* 62 (2), pp. 275–306.
- Stallings, J. A. and G. G. Mohlman (1990). "Issues in qualitative education research: Observation techniques." In H. J. Walberg & G. D. Haertel (Eds.) *The international encyclopedia of educational evaluation* (pp. 639-644). New York, NY: Pergamon Press.
- United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2019). "Global education monitoring report, 2019: Migration, displacement and education: building bridges, not walls." Paris, France: UNESCO.
- United Nations Educational, Scientific, and Cultural Organization (UNESCO)/World Health Organization (WHO)/World Bank (2020). "UNICEF/WHO/World Bank joint child malnutrition estimates (country level)." URL: <https://data.unicef.org/topic/nutrition/malnutrition>. Last accessed: December 29, 2020.
- Vogl, T. (2012). "Education and health in developing economies." *Encyclopedia of Health Economics* 1453, pp. 246–249.
- World Bank (2018). "World Development Report 2018: Learning to realize education's promise." Washington, DC: The World Bank.

Figure 1: Distributional treatment effects on achievement

A. Quantile treatment effects



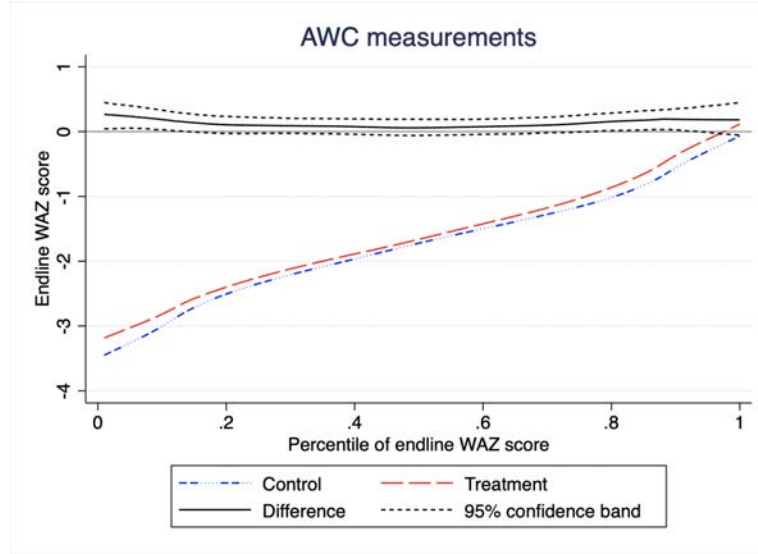
B. Average treatment effects by baseline score



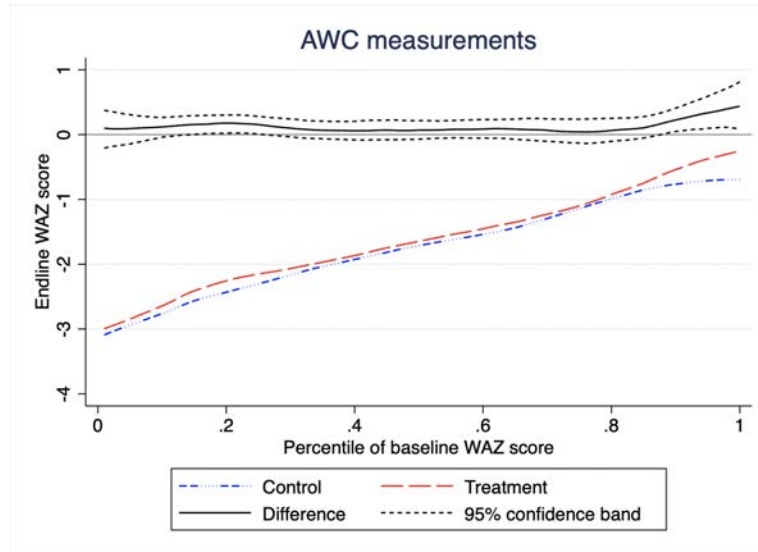
Notes: Panel A shows quantiles of endline composite test scores for treated and control children who participated in the baseline and endline AWC assessments, estimated by local polynomial regressions of endline scores on endline percentiles separately by experimental group. The solid black line plots 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 2: Distributional treatment effects on WAZ scores

A. Quantile treatment effects



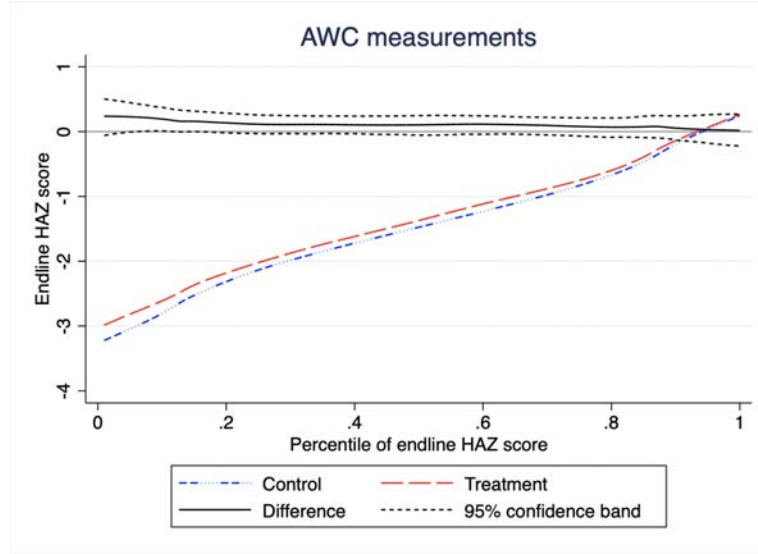
B. Average treatment effects by baseline score



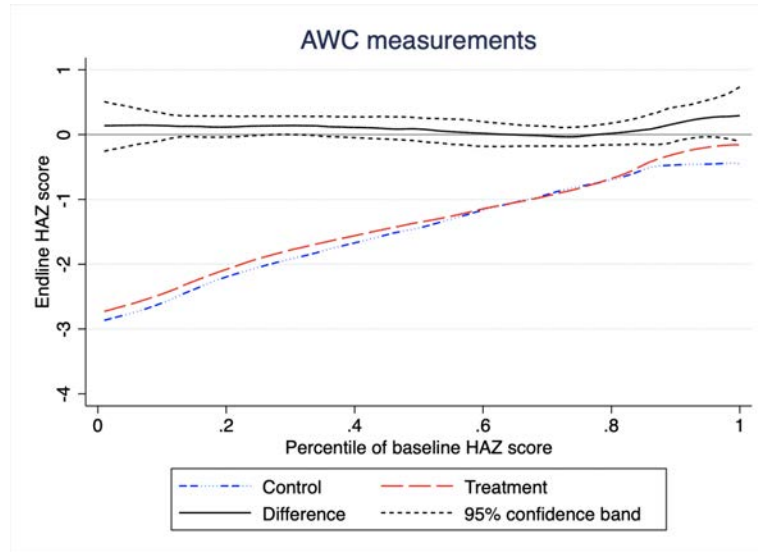
Notes: Panel A shows quantiles of endline weight-for-age z-scores (WAZ) for treated and control children who participated in the baseline and endline AWC assessments, estimated by local polynomial regressions of endline scores on endline percentiles separately by experimental group. The solid black line 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 3: Distributional treatment effects on HAZ scores

A. Quantile treatment effects



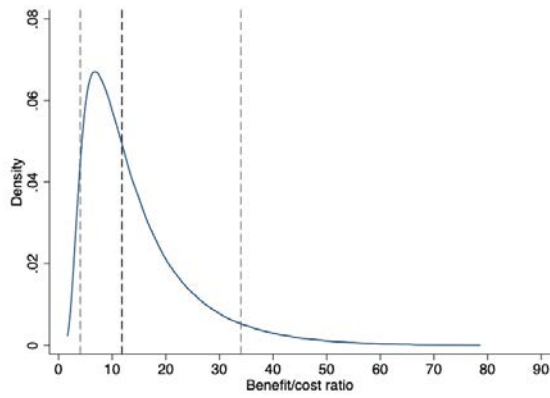
B. Average treatment effects by baseline score



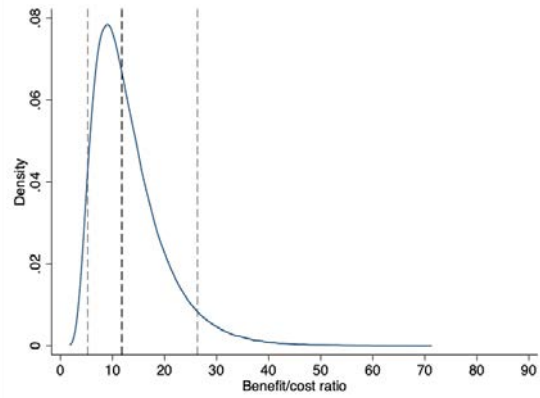
Notes: Panel A shows quantiles of endline height-for-age z-scores (HAZ) for treated and control children who participated in the baseline and endline AWC assessments, estimated by local polynomial regressions of endline scores on endline percentiles separately by experimental group. The solid black line 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 4: Benefit/cost sensitivity analysis

A. Projections based on test score gains for the household sample

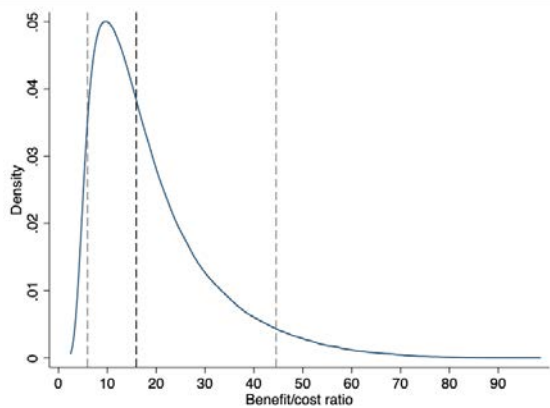


Uniform parameters (mean = 14.2, median = 11.4)

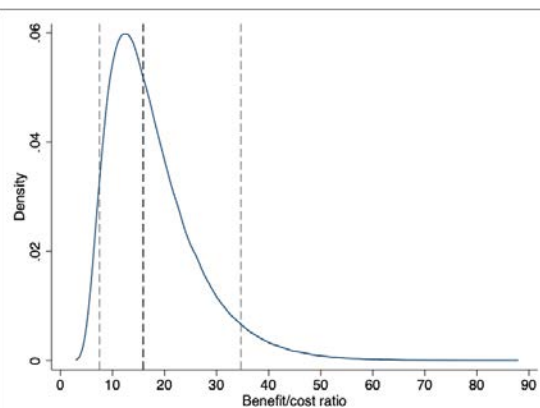


Truncated normal parameters (mean = 13.2, median = 11.6)

B. Projections based on test score and HAZ effects for the AWC sample



Uniform parameters (mean = 19.1, median = 15.7)



Truncated normal parameters (mean = 17.8, median = 15.8)

Notes: This figure explores the sensitivity of the benefit/cost ratio to parameter values. 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 8 in the middle of each range. Days worked in the labor market range from 225 to 275. The wage growth rate ranges from 3 percent to 7 percent. The discount rate ranges from 1.5 percent to 4.5 percent. The proportionate increase in earnings per standard deviation of test scores ranges from 7 percent to 19 percent. The proportionate increase in earnings per standard deviation of HAZ ranges from 14 percent to 26 percent. Panel A displays results based on test score gains for the household sample, corresponding to column 1 of Table 8. Panel B shows results based on test score and HAZ gains for the AWC sample, corresponding to column 3 of Table 8. In each panel the left-hand plot draws each parameter 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. Gray lines indicate 5th and 95th percentiles, and black vertical lines denote our preferred estimates from Table 8.

Table 1: Summary statistics at baseline and randomization balance

	(1) Control	(2) Treatment	(3) Difference	(4) N
<i>A. Anganwadi centers</i>				
No. of children at the AWC	15.006 [6.428]	14.212 [5.384]	-0.794 (0.658)	320
Located next to primary school	0.294 [0.457]	0.312 [0.465]	0.019 (0.051)	320
Functions jointly with another AWC	0.044 [0.205]	0.087 [0.283]	0.044 (0.027)	320
Has electricity connection	0.769 [0.423]	0.762 [0.427]	-0.006 (0.046)	320
Has kitchen	0.856 [0.352]	0.875 [0.332]	0.019 (0.037)	320
Has toilet	0.688 [0.465]	0.762 [0.427]	0.075 (0.050)	320
Has blackboard	0.806 [0.396]	0.769 [0.423]	-0.038 (0.045)	320
<i>B. Anganwadi workers</i>				
Age	50.566 [8.151]	48.481 [9.494]	-2.178** (0.963)	315
Passed grade 10 or higher	0.893 [0.311]	0.966 [0.181]	0.077*** (0.029)	297
Years of experience as AWW	23.094 [9.811]	20.475 [10.696]	-2.619** (1.118)	320
No. days of training last year	6.019 [6.856]	5.308 [4.720]	-0.741 (0.643)	315
Received ECE training	0.606 [0.490]	0.663 [0.474]	0.056 (0.053)	320
Salary (INR)	8,169.394 [1,162.451]	7,861.569 [1,239.866]	-307.825** (128.820)	320
<i>C. Children</i>				
Female	0.504 [0.500]	0.518 [0.500]	0.013 (0.013)	4,675
Age	3.559 [0.848]	3.468 [0.868]	-0.082** (0.033)	4,661
Weight-for-age (WAZ) score	-1.658 [1.025]	-1.599 [1.017]	0.046 (0.039)	4,568
Height-for-age (HAZ) score	-1.565 [1.470]	-1.483 [1.358]	0.071 (0.051)	4,528
Severely or moderately underweight	0.369 [0.483]	0.350 [0.477]	-0.016 (0.017)	4,568
Severely or moderately stunted	0.352 [0.478]	0.334 [0.472]	-0.015 (0.016)	4,528
Math (std. proportion-correct score)	-0.021 [0.994]	0.022 [1.006]	0.015 (0.046)	4,675
Language (std. proportion-correct score)	-0.011 [1.001]	0.011 [0.999]	-0.012 (0.045)	4,675
Exec. function (std. proportion-correct score)	-0.004 [0.984]	0.004 [1.016]	-0.007 (0.045)	4,675
Composite score	-0.022 [1.433]	0.023 [1.448]	-0.001 (0.068)	4,675

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
<i>A. Follow-up rate</i>		
Treatment	-0.008 (0.016)	-0.022 (0.018)
N (children)	4675	2336
Control mean	0.328	0.892
<i>B. Follow-up rate by baseline characteristics</i>		
Treatment	-0.044 (0.075)	-0.053 (0.080)
Female	0.017 (0.019)	-0.014 (0.020)
Age (at baseline)	-0.203*** (0.013)	-0.011 (0.015)
WAZ score (at baseline)	-0.055*** (0.013)	0.050*** (0.017)
HAZ score (at baseline)	-0.013 (0.009)	-0.064*** (0.018)
Composite score (at baseline)	-0.015** (0.007)	0.019* (0.011)
Female \times Treatment	-0.037 (0.026)	-0.007 (0.031)
Age \times Treatment	0.009 (0.018)	-0.001 (0.022)
WAZ \times Treatment	0.001 (0.020)	-0.039 (0.025)
HAZ \times Treatment	-0.008 (0.015)	0.024 (0.024)
Composite score \times Treatment	0.002 (0.010)	0.007 (0.014)
N (children)	4521	1948
F-ratio (Interactions)	0.602	0.552
P-value	0.698	0.737

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	
<i>A. Center-level impacts</i>	Control	Treatment	Col. (2)-(1)	
Share of centers that were...				
...open by AWC opening time (9am)	0.400 [0.493]	0.488 [0.503]	0.049 (0.081)	
...open by PSE start time (10am)	0.875 [0.333]	0.962 [0.191]	0.089** (0.045)	
...closed	0.125 [0.333]	0.038 [0.191]	-0.089** (0.045)	
N (centers)	80	80	160	
	(1)	(2)	(3)	(4)
	AWWs		Impact on AWWs	ECE facilitators
<i>B. Worker-level impacts</i>	Control	Treatment	Col. (2)-(1)	Treatment
Share of workers who...				
...arrived by AWC opening time (9am)	0.125 [0.333]	0.237 [0.428]	0.100 (0.064)	0.213 [0.412]
...arrived by PSE start time (10am)	0.800 [0.403]	0.900 [0.302]	0.107* (0.059)	0.913 [0.284]
...were absent	0.200 [0.403]	0.100 [0.302]	-0.107* (0.059)	0.087 [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) Treatment	(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 [31.530]	38.400 [29.665]	18.150 [21.432]	-19.014*** (4.183)	75.600 [37.092]	38.908*** (5.386)
...on administrative tasks	19.650 [17.834]	21.900 [22.084]	35.100 [26.715]	13.091*** (4.022)	54.750 [34.519]	32.780*** (4.712)
...on health and nutrition tasks	5.550 [8.527]	5.700 [9.917]	10.800 [14.616]	5.691*** (1.977)	16.350 [18.338]	11.317*** (2.315)
...off duty	26.850 [24.019]	30.000 [26.799]	43.950 [30.366]	13.131*** (4.743)	70.800 [45.808]	39.751*** (6.045)
...absent	10.500 [34.122]	24.000 [48.303]	12.000 [36.227]	-12.899* (7.094)	22.500 [60.661]	-2.755 (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 pre-school 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)	(2)	(3)	(4)
	Math	Language	Executive function	Composite score
<i>A. Complete sample</i>				
<u>AWC assessments</u>				
Treatment	0.291*** (0.061)	0.459*** (0.081)	0.180*** (0.052)	0.288*** (0.058)
N (children)	1514	1514	1514	1514
<u>HH assessments</u>				
Treatment	0.128*** (0.049)	0.102** (0.051)	0.054 (0.042)	0.109** (0.045)
N (children)	2080	2080	2080	2080
<i>B. Common sample</i>				
<u>AWC assessments</u>				
Treatment	0.311*** (0.075)	0.460*** (0.095)	0.205*** (0.068)	0.315*** (0.071)
N (children)	791	791	791	791
<u>HH assessments</u>				
Treatment	0.290*** (0.080)	0.361*** (0.091)	0.158** (0.062)	0.291*** (0.071)
N (children)	791	791	791	791
P-value (AWC = HH)	0.935	0.330	0.407	0.727

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 dependant 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 scores

	(1)	(2)	(3)
	WAZ score	Underweight (WAZ<-2)	Severely underweight (WAZ<-3)
<i>A. Complete sample</i>			
<u>AWC measurements</u>			
Treatment	0.096*** (0.033)	-0.018 (0.017)	-0.031** (0.012)
N (children)	1538	1538	1538
Control mean	-1.762	0.384	0.091
<u>HH measurements</u>			
Treatment	0.047 (0.032)	-0.013 (0.018)	-0.007 (0.010)
N (children)	2021	2021	2021
Control mean	-1.553	0.321	0.075
<i>B. Common sample</i>			
<u>AWC measurements</u>			
Treatment	0.119*** (0.037)	-0.021 (0.025)	-0.045** (0.020)
N (children)	791	791	791
Control mean	-2.001	0.517	0.138
<u>HH measurements</u>			
Treatment	0.074* (0.041)	-0.007 (0.030)	-0.021 (0.021)
N (children)	791	791	791
Control mean	-2.001	0.515	0.153
P-value (AWC = HH)	0.179	0.573	0.134

Notes: The table shows the impact of the intervention on children's weight-for-age z -scores (WAZ), the share of underweight children (WAZ below -2), and the share of severely underweight children (WAZ below -3). Estimates come from regressions of WAZ 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 dependant 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 7: Impact on endline HAZ scores

	(1)	(2)	(3)
	HAZ score	Stunted (HAZ<-2)	Severely stunted (HAZ<-3)
<i>A. Complete sample</i>			
<u>AWC measurements</u>			
Treatment	0.092** (0.044)	-0.048** (0.022)	-0.023** (0.011)
N (children)	1389	1389	1389
Control mean	-1.492	0.291	0.057
<u>HH measurements</u>			
Treatment	0.011 (0.054)	-0.027 (0.017)	-0.011 (0.007)
N (children)	1995	1995	1995
Control mean	-1.164	0.205	0.040
<i>B. Common sample</i>			
<u>AWC measurements</u>			
Treatment	0.153*** (0.054)	-0.090*** (0.028)	-0.042*** (0.016)
N (children)	724	724	724
Control mean	-1.674	0.375	0.080
<u>HH measurements</u>			
Treatment	0.101* (0.055)	-0.053* (0.030)	-0.026 (0.016)
N (children)	724	724	724
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 height-for-age z-scores (HAZ), the share of stunted children (HAZ below -2), and the share of severely stunted children (HAZ below -3). Estimates come from regressions of 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 dependant 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 8: Cost/benefit analysis

Parameter	Source	(1) HH sample, test scores	(2) AWC sample, test scores	(3) AWC sample, tests + HAZ	(4) AWC sample, tests + stunting
<i>A. Projecting future earnings</i>					
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	
Discount rate	Assumption			0.03	
Average PDV of lifetime earnings	Calculation			3,311,001	
<i>B. Experimental impacts</i>					
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	47,347	124,825	178,999	226,125
<i>C. Benefit/cost ratio</i>					
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	881,607	830,085	1,190,346	1,503,730
Program cost per center	Government order	74,478	74,478	74,478	74,478
Benefit/cost ratio	Calculation	11.8	11.1	16.0	20.2

Notes: This table reports a cost benefit analysis of the early-childhood education 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 rural Tamil Nadu. Panel B lists parameters and assumptions necessary to predict the increase in earnings generated by the intervention for each child. Panel C combines this projection with program costs to produce a benefit/cost ratio. Column 1 uses the full household sample and projects benefits using impacts on test scores. Column 2 assumes all benefits accrue to children in the AWC sample and projects benefits using impacts on test scores. Column 3 combines projections based on test scores and HAZ scores for the AWC sample. Column 4 combines projections based on test scores and stunting for 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 Table 7, Panel A, columns 1 and 2.

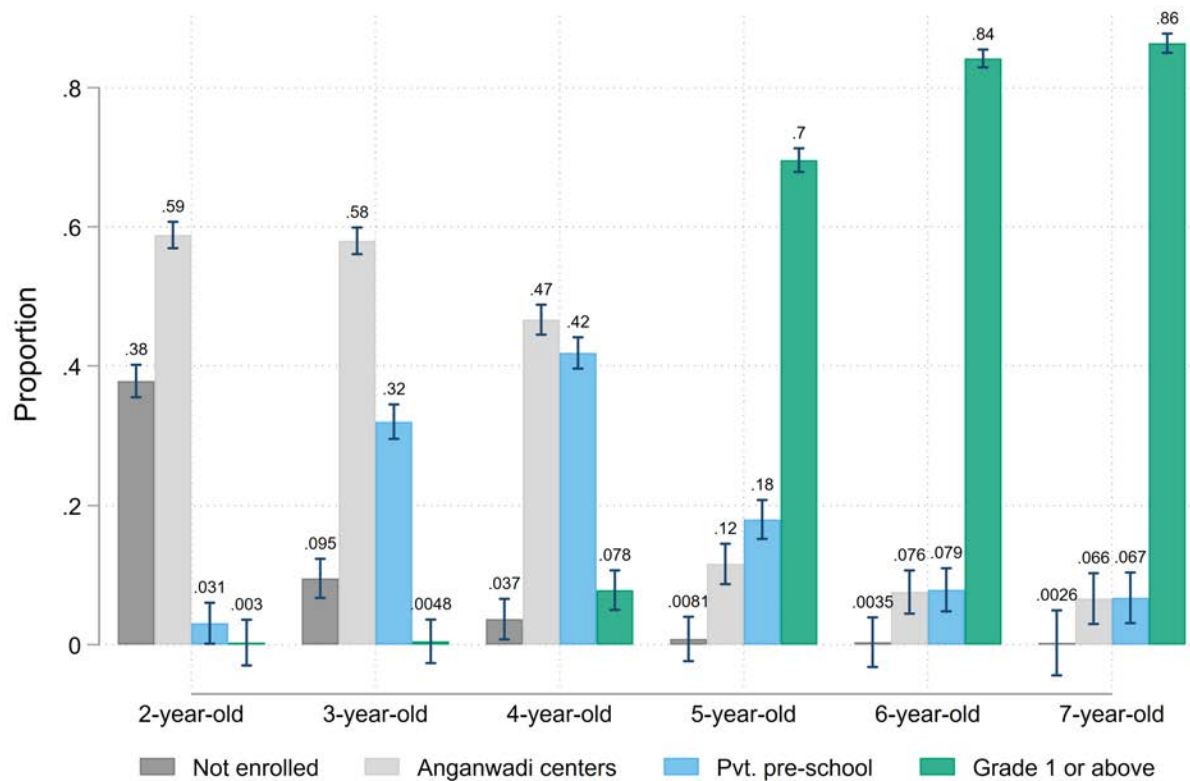
^g Hoddinott et al. (2011) estimate that a standard deviation increase in HAZ increases adult consumption by 20% and that stunting reduces adult consumption by 66%.

^h Table A.20, 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%.

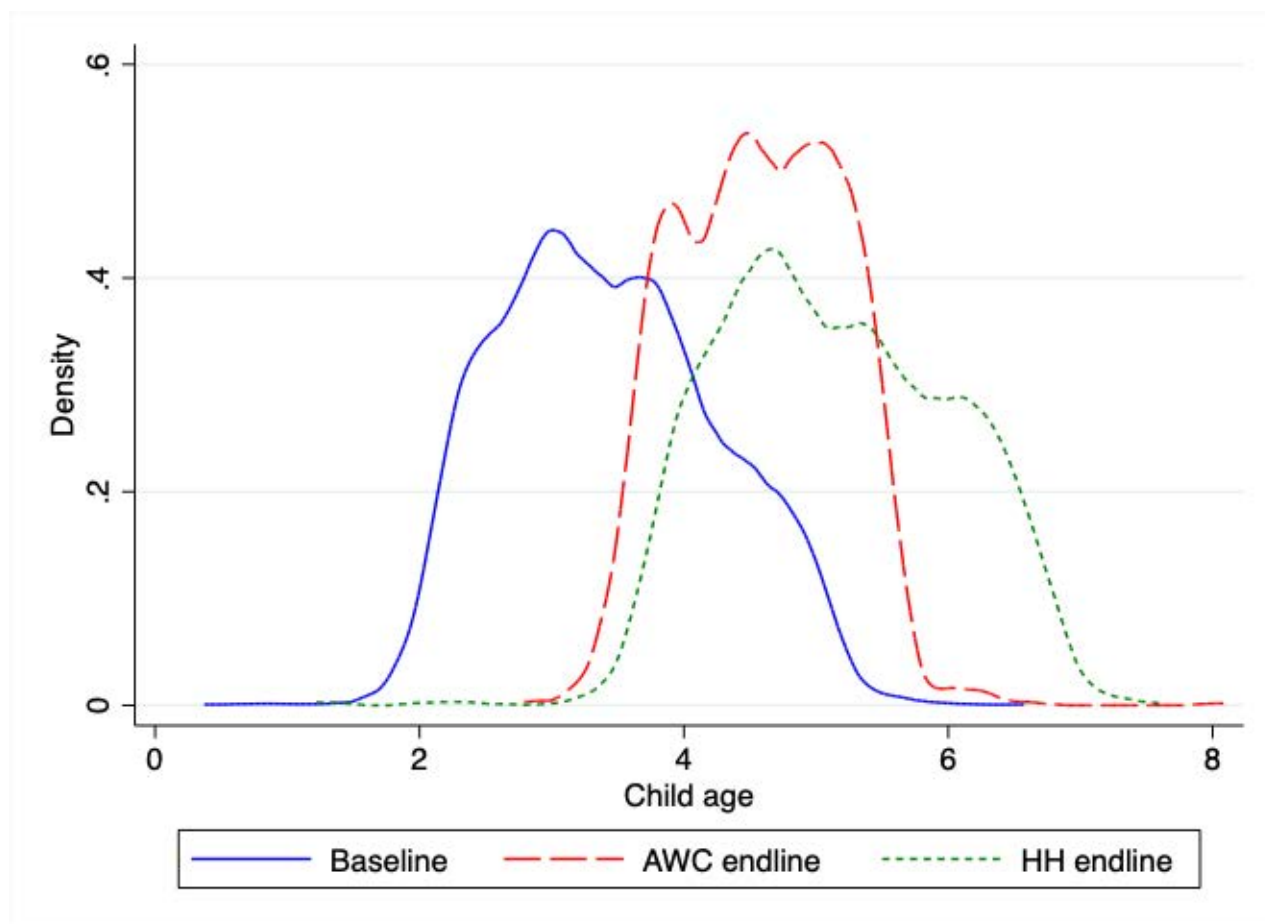
Appendix A Additional figures and tables

Figure A.1: 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 pre-schooling options and the sample comprises 23,717 children (aged 2-7) across 197 villages.

Figure A.2: 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.

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) ECE facilitators	(2) AWWs	(3) AWWs	(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)
...preparing or serving food	2.850 [6.681]	1.800 [5.088]	5.850 [12.664]	4.622*** (1.506)	8.700 [15.509]	7.552*** (1.786)
...assisting children to use toilet	2.400 [4.830]	2.550 [4.940]	1.650 [4.159]	-0.846 (0.756)	4.050 [5.710]	1.562* (0.873)
...on health-related activities	0.300 [1.885]	1.350 [6.331]	3.300 [6.607]	1.915* (1.035)	3.600 [7.245]	2.202** (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 pre-school 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 time allocation to administrative tasks from unannounced observations

	(1) ECE facilitators	(2)	(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)
...cleaning or organizing center	1.050 [3.910]	0.300 [1.885]	1.800 [5.760]	1.500** (0.705)	2.850 [6.948]	2.540*** (0.828)
...completing registers	0.450 [2.294]	6.900 [16.288]	20.550 [24.836]	13.250*** (3.382)	21.000 [25.150]	13.669*** (3.409)
...on classroom management	18.150 [17.183]	14.700 [15.272]	12.750 [14.954]	-1.658 (2.486)	30.900 [25.884]	16.571*** (3.397)
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 pre-school 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.4: Impact on time off task from unannounced observations

	(1) ECE facilitators	(2) AWWs	(3) AWWs	(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)
...uninvolved	10.950 [16.335]	9.900 [13.058]	18.450 [20.020]	8.209*** (2.713)	29.400 [28.120]	18.897*** (3.484)
...out of center (reasons unknown)	2.700 [6.325]	3.000 [7.271]	7.500 [16.243]	4.471** (1.961)	10.200 [17.819]	7.068*** (2.122)
...out of center (adm. duties)	0.000 [0.000]	0.150 [1.342]	1.350 [6.612]	1.204 (0.755)	1.350 [6.612]	1.204 (0.755)
...out of center (home visitations)	0.000 [0.000]	0.000 [0.000]	0.300 [1.885]	0.296 (0.210)	0.300 [1.885]	0.296 (0.210)
...out of center (health duties)	0.000 [0.000]	1.050 [9.391]	1.650 [11.415]	0.483 (1.645)	1.650 [11.415]	0.483 (1.645)
...on social interaction	12.750 [12.570]	10.500 [12.861]	11.700 [13.632]	1.735 (2.126)	24.450 [19.040]	14.598*** (2.588)
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 pre-school 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.5: Facilitator self-reports on time allocation from intervention monitoring

Hours spent last week...	(1)
<i>A. Education</i>	
...teaching pre-school education	17.452 [9.198]
<i>B. Health and nutrition</i>	
...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]
<i>C. Administrative</i>	
...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 pre-school 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 A.6: Impact on allocation of time on task from announced observations

	(1) ECE facilitators	(2) AWWs	(3) AWWs	(4) Impact on AWWs	(5) AWWs & facilitators Col. (1)+(3)	(6) Same time	(7) Impact on AWCs
Minutes per day...	Treatment	Control	Treatment	Col. (3)-(2)	Overall		Col. (5)-(2)
...reading aloud	0.300 [1.897]	0.900 [3.201]	0.000 [0.000]	-0.936* (0.537)	0.300 [1.897]	0.000 [0.000]	-0.608 (0.630)
...on demonstration	0.600 [2.649]	0.600 [2.649]	0.300 [1.897]	-0.304 (0.543)	0.900 [3.201]	0.000 [0.000]	0.319 (0.690)
...on questions and answers	10.800 [10.802]	6.600 [8.130]	2.100 [4.618]	-4.545*** (1.538)	12.900 [11.965]	0.000 [0.000]	5.887** (2.344)
...on practice and drill	30.600 [21.215]	24.900 [18.105]	7.800 [11.055]	-17.352*** (3.453)	38.400 [23.721]	4.200 [8.827]	14.742*** (4.720)
...on classwork	18.300 [15.127]	18.900 [16.727]	11.100 [16.165]	-7.497* (3.867)	29.400 [25.046]	3.300 [9.008]	10.418** (5.020)
N (centers)	40	40	40	80	40	40	80

Notes: This table compares the instructional time allocation of AWWs in control and treatment AWCs and of AWWs and facilitators in treatment AWCs, measured in announced 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 pre-school 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.7: Impacts on parent-reported home visitation from household survey

	(1) At least one visit in last 3 months	(2) Three or more visits in last 3 months	(3) At least one visit in last 6 months	(4) Three or more visits in last 6 months
Treatment	0.014 (0.029)	-0.006 (0.022)	0.014 (0.029)	0.018 (0.027)
N (children)	1950	1950	1940	1940
Control mean	0.408	0.174	0.425	0.318

Notes: The table shows the impact of the intervention on parent-reported home visitations. Estimates come from regressions of each variable on a treatment indicator, with controls for randomization strata. AWW experience and education, and missingness, for children who were present in the baseline and the household endline assessments. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.8: Impact on endline assessments (proportion-correct scores)

	(1)	(2)	(3)
	Math	Language	Executive function
<i>A. Complete sample</i>			
<u>AWC assessments</u>			
Treatment	0.052*** (0.011)	0.074*** (0.013)	0.055*** (0.016)
N (children)	1514	1514	1514
Control mean	0.194	0.122	0.382
<u>HH assessments</u>			
Treatment	0.041*** (0.016)	0.033** (0.017)	0.014 (0.011)
N (children)	2080	2080	2080
Control mean	0.338	0.312	0.412
<i>B. Common sample</i>			
<u>AWC assessments</u>			
Treatment	0.055*** (0.013)	0.074*** (0.015)	0.062*** (0.021)
N (children)	791	791	791
Control mean	0.203	0.128	0.404
<u>HH assessments</u>			
Treatment	0.061*** (0.017)	0.077*** (0.019)	0.039** (0.015)
N (children)	791	791	791
Control mean	0.208	0.186	0.359
P-value (AWC = HH)	0.621	0.763	0.207

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 dependant 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 A.9: 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.
<i>A. AWC</i>							
Treatment	0.078*** (0.016)	0.416*** (0.088)	0.078*** (0.018)	0.462*** (0.106)	0.074*** (0.023)	0.227*** (0.071)	0.373*** (0.079)
Baseline score	0.399*** (0.096)	0.384*** (0.093)	0.296*** (0.100)	0.333*** (0.113)	0.283*** (0.049)	0.219*** (0.038)	0.285*** (0.055)
N (children)	791	791	791	791	791	791	791
Control mean	0.148	0.000	0.117	0.000	0.403	0.000	-0.000
<i>B. Household</i>							
Treatment	0.061*** (0.017)	0.290*** (0.080)	0.077*** (0.019)	0.361*** (0.091)	0.039** (0.015)	0.158** (0.062)	0.291*** (0.071)
Baseline score	0.461*** (0.102)	0.395*** (0.088)	0.313*** (0.098)	0.280*** (0.087)	0.206*** (0.035)	0.214*** (0.037)	0.282*** (0.054)
N (children)	791	791	791	791	791	791	791
Control mean	0.208	-0.000	0.186	0.000	0.359	-0.000	0.000
P-value ($T \times AWC$)	0.576	0.338	0.635	0.145	0.267	0.554	0.422

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, baseline child and center scores, AWW experience and education, and missingness. Raw scores are expressed as proportions of questions in each test answered correctly; standardized scores are rescaled to have a mean of zero and a standard deviation of one. Both panels display results for children who participated in the baseline and both endline assessments for the overlapping items across both rounds. Panels A and B display results for the AWC- and HH-based assessments, respectively. None of the estimates use weights. All specifications control for a baseline measure of the dependant variable, AWW experience, and AWW education. The last row displays the p-value on the interaction term between treatment and an indicator variable for AWC assessments. Standard errors (clustered by AWC) appear in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.10: Impact on endline assessments by AWC attendance (standardized scores)

	(1)	(2)	(3)	(4)
	Math	Language	Executive function	Composite score
<i>A. Attends AWC</i>				
<u>HH assessments</u>				
Treatment	0.154*** (0.057)	0.221*** (0.056)	0.127** (0.055)	0.185*** (0.052)
N (children)	1132	1132	1132	1132
Control mean	-0.434	-0.418	-0.259	-0.402
<i>B. Does not attend</i>				
<u>HH assessments</u>				
Treatment	0.060 (0.062)	-0.066 (0.067)	-0.032 (0.057)	-0.007 (0.060)
N (children)	948	948	948	948
Control mean	0.828	0.784	0.467	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 dependant 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.11: 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
<i>A. Complete sample</i>							
<u>AWC measurements</u>							
Treatment	0.263** (0.119)	0.201*** (0.072)	0.127 (0.264)	0.343*** (0.092)	0.271*** (0.064)	0.401 (0.432)	0.267*** (0.065)
Covariate	0.011 (0.038)	0.010 (0.066)	0.043*** (0.005)	0.110 (0.071)	0.050 (0.103)	0.049 (0.098)	0.409*** (0.051)
Interaction	-0.013 (0.054)	0.172* (0.097)	0.005 (0.007)	-0.069 (0.104)	0.055 (0.146)	-0.040 (0.151)	-0.036 (0.071)
N (children)	1511	1514	1511	1514	1514	1514	1514
<u>HH measurements</u>							
Treatment	0.062 (0.090)	-0.075 (0.058)	0.289 (0.196)	0.099* (0.059)	0.103* (0.054)	0.256 (0.302)	0.092** (0.045)
Covariate	0.074** (0.031)	-0.073 (0.059)	0.042*** (0.003)	0.012 (0.057)	0.111 (0.095)	0.008 (0.088)	0.420*** (0.035)
Interaction	-0.028 (0.044)	0.359*** (0.084)	-0.004 (0.004)	0.022 (0.079)	-0.005 (0.112)	-0.054 (0.110)	0.001 (0.043)
N (children)	2080	2080	2080	2080	2080	2080	2080
<i>B. Common sample</i>							
<u>AWC measurements</u>							
Treatment	0.197 (0.159)	0.237** (0.093)	-0.054 (0.297)	0.364*** (0.096)	0.280*** (0.084)	0.496 (0.540)	0.299*** (0.080)
Covariate	0.062 (0.049)	0.094 (0.081)	0.042*** (0.006)	0.124 (0.088)	-0.072 (0.128)	-0.085 (0.128)	0.403*** (0.075)
Interaction	-0.056 (0.070)	0.141 (0.127)	0.010 (0.008)	-0.087 (0.133)	0.147 (0.159)	-0.068 (0.191)	-0.025 (0.094)
<u>HH measurements</u>							
Treatment	0.315* (0.177)	0.261*** (0.099)	-0.279 (0.357)	0.208** (0.095)	0.277*** (0.078)	-0.564 (0.642)	0.299*** (0.080)
Covariate	0.052 (0.055)	0.163* (0.093)	0.032*** (0.007)	-0.024 (0.094)	-0.052 (0.174)	-0.130 (0.140)	0.403*** (0.075)
Interaction	0.015 (0.078)	0.047 (0.146)	0.016 (0.010)	0.194 (0.141)	0.061 (0.210)	0.305 (0.224)	-0.025 (0.094)
N (children)	791	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.12: Heterogeneous impacts on endline WAZ scores

	(1) Baseline WAZ	(2) Female	(3) Age (months)	(4) Mother education	(5) AWW vacancy	(6) Class size	(7) Baseline score
<i>A. Complete sample</i>							
<u>AWC measurements</u>							
Treatment	0.323*** (0.115)	0.085* (0.044)	0.326 (0.235)	0.059 (0.050)	0.100*** (0.037)	0.190 (0.276)	0.102** (0.042)
Covariate	0.693*** (0.046)	-0.078* (0.042)	0.005 (0.004)	0.022 (0.051)	0.007 (0.063)	0.041 (0.066)	-0.063** (0.030)
Interaction	0.128** (0.055)	0.021 (0.058)	-0.006 (0.006)	0.056 (0.062)	-0.015 (0.087)	-0.033 (0.096)	0.028 (0.040)
N (children)	1538	1538	1538	1538	1538	1538	1509
<u>HH measurements</u>							
Treatment	0.151 (0.094)	0.046 (0.044)	0.114 (0.136)	0.055 (0.041)	0.054 (0.035)	0.195 (0.248)	0.055* (0.032)
Covariate	0.779*** (0.034)	0.068 (0.052)	-0.001 (0.002)	0.028 (0.044)	0.111 (0.074)	-0.068 (0.061)	-0.001 (0.023)
Interaction	0.066 (0.050)	-0.002 (0.070)	-0.002 (0.003)	-0.014 (0.065)	-0.048 (0.073)	-0.056 (0.087)	0.023 (0.027)
N (children)	2021	2021	2021	2021	2021	2021	1989
<i>B. Common sample</i>							
<u>AWC measurements</u>							
Treatment	0.416*** (0.120)	0.109** (0.052)	0.163 (0.227)	0.080* (0.048)	0.128*** (0.042)	0.534* (0.302)	0.143*** (0.043)
Covariate	0.729*** (0.042)	-0.062 (0.055)	-0.001 (0.005)	-0.029 (0.063)	0.108 (0.079)	0.040 (0.069)	-0.047 (0.035)
Interaction	0.145*** (0.052)	0.022 (0.074)	-0.001 (0.006)	0.087 (0.078)	-0.055 (0.094)	-0.148 (0.106)	0.047 (0.044)
<u>HH measurements</u>							
Treatment	0.264** (0.120)	0.046 (0.062)	-0.025 (0.249)	0.047 (0.053)	0.047 (0.045)	0.340 (0.313)	0.099** (0.046)
Covariate	0.767*** (0.048)	0.016 (0.062)	-0.003 (0.005)	0.032 (0.055)	0.054 (0.074)	0.137* (0.081)	-0.037 (0.035)
Interaction	0.105* (0.060)	0.051 (0.083)	0.003 (0.006)	0.071 (0.078)	0.087 (0.094)	-0.092 (0.113)	0.049 (0.043)
N (children)	791	791	791	791	791	791	780

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.13: Heterogeneous impacts on endline HAZ scores

	(1) Baseline HAZ	(2) Female	(3) Age (months)	(4) Mother education	(5) AWW vacancy	(6) Class size	(7) Baseline score
<i>A. Complete sample</i>							
<u>AWC measurements</u>							
Treatment	0.316* (0.178)	0.086 (0.059)	0.286 (0.294)	0.148** (0.069)	0.074 (0.051)	-0.109 (0.398)	0.111** (0.049)
Covariate	0.481*** (0.069)	-0.006 (0.053)	0.009* (0.005)	0.152** (0.064)	-0.019 (0.085)	-0.025 (0.085)	-0.066* (0.034)
Interaction	0.130 (0.090)	0.011 (0.075)	-0.005 (0.008)	-0.069 (0.079)	0.074 (0.112)	0.072 (0.139)	0.058 (0.040)
N (children)	1389	1389	1389	1389	1389	1389	1362
<u>HH measurements</u>							
Treatment	-0.211 (0.210)	-0.016 (0.078)	-0.355 (0.329)	0.113* (0.063)	0.080 (0.051)	0.065 (0.361)	0.020 (0.053)
Covariate	0.611*** (0.053)	-0.082 (0.074)	-0.002 (0.005)	0.145** (0.064)	0.383** (0.162)	-0.018 (0.083)	-0.018 (0.030)
Interaction	-0.151 (0.125)	0.056 (0.097)	0.009 (0.007)	-0.189* (0.096)	-0.317** (0.148)	-0.021 (0.128)	0.089*** (0.033)
N (children)	1995	1995	1995	1995	1995	1995	1963
<i>B. Common sample</i>							
<u>AWC measurements</u>							
Treatment	0.409** (0.196)	0.149** (0.073)	0.666** (0.309)	0.155** (0.067)	0.143** (0.057)	0.234 (0.418)	0.159*** (0.055)
Covariate	0.553*** (0.057)	0.029 (0.068)	0.011** (0.005)	0.044 (0.072)	0.058 (0.121)	0.036 (0.099)	0.004 (0.036)
Interaction	0.133 (0.089)	0.005 (0.096)	-0.014* (0.008)	0.004 (0.097)	0.025 (0.134)	-0.028 (0.147)	0.001 (0.044)
<u>HH measurements</u>							
Treatment	0.347* (0.197)	0.098 (0.074)	0.941* (0.483)	0.154** (0.077)	0.086 (0.057)	1.019** (0.438)	0.123** (0.058)
Covariate	0.523*** (0.066)	0.021 (0.090)	0.024** (0.012)	0.199** (0.092)	0.010 (0.124)	0.187* (0.109)	-0.003 (0.053)
Interaction	0.132 (0.104)	0.005 (0.115)	-0.022* (0.013)	-0.082 (0.130)	0.053 (0.145)	-0.325** (0.154)	0.030 (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.14: Impacts on WAZ scores after dropping child outliers based on residuals

	(1)	(2)	(3)
	WAZ score	Underweight (WAZ<-2)	Severely underweight (WAZ<-3)
<i>A. Dropping residuals > 0.5</i>			
Treatment	0.086** (0.034)	-0.023 (0.021)	-0.030*** (0.010)
N (children)	1355	1355	1355
Control mean	-1.898	0.427	0.103
<i>B. Dropping residuals > 0.4</i>			
Treatment	0.083** (0.035)	-0.026 (0.021)	-0.033*** (0.011)
N (children)	1289	1289	1289
Control mean	-1.898	0.427	0.103
<i>C. Dropping residuals > 0.3</i>			
Treatment	0.106*** (0.035)	-0.037 (0.023)	-0.038*** (0.012)
N (children)	1182	1182	1182
Control mean	-1.898	0.427	0.103
<i>D. Dropping residuals > 0.2</i>			
Treatment	0.109*** (0.038)	-0.052** (0.025)	-0.042*** (0.014)
N (children)	1060	1060	1060
Control mean	-1.898	0.427	0.103
<i>E. Dropping residuals > 0.1</i>			
Treatment	0.122*** (0.043)	-0.068** (0.029)	-0.049*** (0.016)
N (children)	915	915	915
Control mean	-1.898	0.427	0.103

Notes: The table replicates the estimation from panel A of Table 6 with different levels of outlier exclusions. A child is marked as an outlier if the absolute value of the residual in the impact estimation model is greater than the given amount. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.15: Impacts on HAZ scores after dropping child outliers based on residuals

	(1)	(2)	(3)
	HAZ score	Stunted (HAZ<-2)	Severely stunted (HAZ<-3)
<i>A. Dropping residuals > 0.5</i>			
Treatment	0.122*** (0.044)	-0.058** (0.026)	-0.021* (0.011)
N (children)	1131	1131	1131
Control mean	-1.782	0.361	0.070
<i>B. Dropping residuals > 0.4</i>			
Treatment	0.148*** (0.043)	-0.078*** (0.028)	-0.025** (0.012)
N (children)	1043	1043	1043
Control mean	-1.782	0.361	0.070
<i>C. Dropping residuals > 0.3</i>			
Treatment	0.162*** (0.044)	-0.091*** (0.030)	-0.028** (0.013)
N (children)	959	959	959
Control mean	-1.782	0.361	0.070
<i>D. Dropping residuals > 0.2</i>			
Treatment	0.172*** (0.047)	-0.097*** (0.033)	-0.031** (0.014)
N (children)	865	865	865
Control mean	-1.782	0.361	0.070
<i>E. Dropping residuals > 0.1</i>			
Treatment	0.176*** (0.053)	-0.118*** (0.038)	-0.039** (0.017)
N (children)	752	752	752
Control mean	-1.782	0.361	0.070

Notes: The table replicates the estimation from panel A of Table 7 with different levels of outlier exclusions. A child is marked as an outlier if the absolute value of the residual in the impact estimation model is greater than the given amount. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.16: Impacts on winsorized WAZ scores

	(1) WAZ score	(2) Underweight (WAZ<-2)	(3) Severely underweight (WAZ<-3)
<i>A. Winsorized at 0.1%</i>			
Treatment	0.103*** (0.034)	-0.021 (0.018)	-0.025*** (0.009)
N (children)	1538	1538	1538
Control mean	-1.756	0.384	0.091
<i>B. Winsorized at 0.2%</i>			
Treatment	0.105*** (0.035)	-0.021 (0.018)	-0.025*** (0.009)
N (children)	1538	1538	1538
Control mean	-1.759	0.384	0.091
<i>C. Winsorized at 0.5%</i>			
Treatment	0.105*** (0.036)	-0.021 (0.018)	-0.025*** (0.009)
N (children)	1538	1538	1538
Control mean	-1.760	0.384	0.091
<i>D. Winsorized at 1%</i>			
Treatment	0.105*** (0.036)	-0.021 (0.018)	-0.025*** (0.009)
N (children)	1538	1538	1538
Control mean	-1.761	0.384	0.091
<i>E. Winsorized at 2%</i>			
Treatment	0.105*** (0.037)	-0.021 (0.018)	-0.025*** (0.009)
N (children)	1538	1538	1538
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 A.17: Impacts on winsorized HAZ scores

	(1)	(2)	(3)
	HAZ score	Stunted (HAZ<-2)	Severely stunted (HAZ<-3)
<i>A. Winsorized at 0.1%</i>			
Treatment	0.080* (0.045)	-0.041** (0.020)	-0.016* (0.008)
N (children)	1389	1389	1389
Control mean	-1.478	0.291	0.057
<i>B. Winsorized at 0.2%</i>			
Treatment	0.081* (0.046)	-0.041** (0.020)	-0.016* (0.008)
N (children)	1389	1389	1389
Control mean	-1.482	0.291	0.057
<i>C. Winsorized at 0.5%</i>			
Treatment	0.083* (0.047)	-0.041** (0.020)	-0.016* (0.008)
N (children)	1389	1389	1389
Control mean	-1.486	0.291	0.057
<i>D. Winsorized at 1%</i>			
Treatment	0.084* (0.048)	-0.041** (0.020)	-0.016* (0.008)
N (children)	1389	1389	1389
Control mean	-1.490	0.291	0.057
<i>E. Winsorized at 2%</i>			
Treatment	0.085* (0.048)	-0.041** (0.020)	-0.016* (0.008)
N (children)	1389	1389	1389
Control mean	-1.491	0.291	0.057

Notes: The table replicates the estimation from panel A of Table 7 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 A.18: Impacts on endline WAZ scores by AWC attendance

	(1)	(2)	(3)
	WAZ score	Underweight (WAZ<-2)	Severely underweight (WAZ<-3)
<i>A. Attends AWC</i>			
<u>HH measurements</u>			
Treatment	0.107*** (0.039)	-0.009 (0.024)	-0.009 (0.014)
N (children)	1097	1097	1097
Control mean	-1.698	0.353	0.084
<i>B. Does not attend</i>			
<u>HH measurements</u>			
Treatment	-0.015 (0.049)	-0.003 (0.025)	-0.001 (0.014)
N (children)	924	924	924
Control mean	-1.386	0.283	0.065

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 dependant 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.19: Impacts on endline HAZ scores by AWC attendance

	(1)	(2)	(3)
	HAZ score	Stunted (HAZ<-2)	Severely stunted (HAZ<-3)
<i>A. Attends AWC</i>			
<u>HH measurements</u>			
Treatment	0.080 (0.056)	-0.042* (0.024)	-0.015 (0.012)
N (children)	1084	1084	1084
Control mean	-1.364	0.240	0.053
<i>B. Does not attend</i>			
<u>HH measurements</u>			
Treatment	-0.014 (0.083)	-0.013 (0.022)	-0.004 (0.009)
N (children)	911	911	911
Control mean	-0.933	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 dependant 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.20: Regression-based association between nutrition status and learning outcomes

	(1)	(2)	(3)	(4)
	Math	Language	Executive function	Composite score
<i>A. Weight-for-age</i>				
Weight-for-age (WAZ) score	0.1***	0.085***	0.07***	0.149***
Underweight (WAZ<-2)	-0.121***	-0.122***	-0.116***	-0.208***
Severely underweight (WAZ<-3)	-0.111***	-0.201***	-0.075***	-0.274***
N (children)	4,568	4,568	4,568	4,568
<i>B. Height-for-age</i>				
Height-for-age (HAZ) score	0.091***	0.081***	0.075***	0.143***
Stunted (HAZ<-2)	-0.189***	-0.214***	-0.235***	-0.365***
Severely stunted (HAZ<-3)	-0.282***	-0.246***	-0.325***	-0.487***
N (children)	4,528	4,528	4,528	4,528

Notes: The table shows the association between 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%.

Appendix B Measurement

B.1 Child assessments

B.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.³⁷

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.³⁸

B.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 B.1. The mean standardized scores are in Table 1.

Figure B.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,

³⁷The baseline assessments can be accessed at: <https://bit.ly/2KGITA9>.

³⁸The 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.2).

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 B.1).

B.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 pre-school education and tracked the amount of time that the worker and facilitator spent on instructional activities for the two hours devoted to pre-school 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.³⁹

B.3 Children’s weight and height

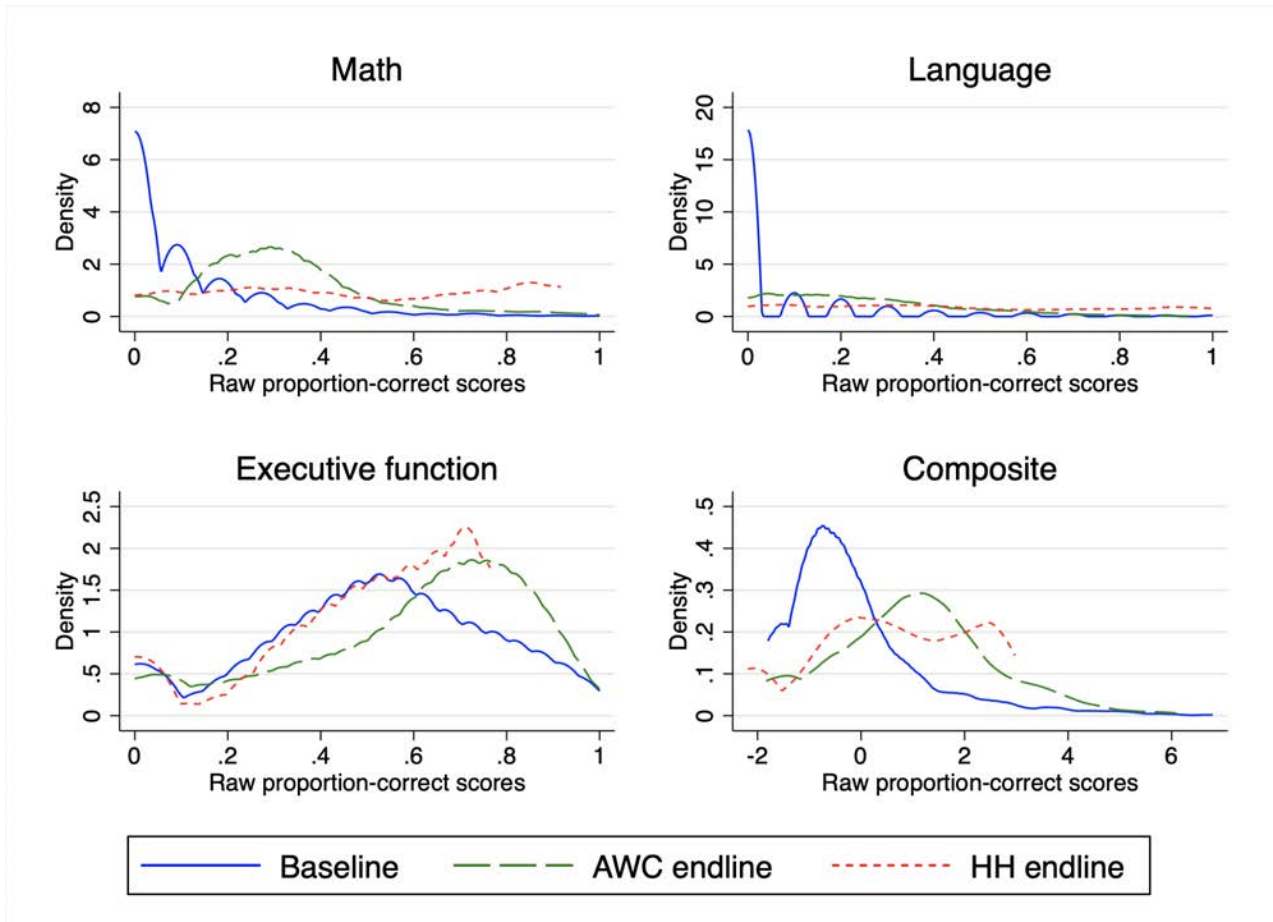
Enumerators from J-PAL South Asia measured each child’s weight as follows. First, they removed the child’s shoes, headpieces, accessories, and jewelry; they checked that the child’s pockets were empty; and made sure that the child was wearing light-weight clothes and bare feet. Then, they placed the 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 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.⁴⁰

³⁹The protocols for the announced visits can be accessed at: <https://bit.ly/2DZri56>.

⁴⁰The protocols for the measurements can be accessed at: <https://bit.ly/2VtT1lF>.

Figure B.1: Raw 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 all children in the study. The proportion-correct score for the composite panel displays the proportion of items answered correctly across all three assessments. The figure includes all children in a given round, irrespective of whether they were present in the other rounds.

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

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

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.