

Teachers on the Move: Evidence From a Large-Scale Learning Intervention During Lockdown

KARTIKA BHATIA* & MARGARET LEIGHTON**

*Centre for Global Learning (GLEA), Coventry University, Coventry, United Kingdom; **Department of Economics, University of St Andrews, Fife, United Kingdom

(Original version submitted October 2022; final version accepted March 2024)

ABSTRACT *The move to remote learning during Covid-19 school closures left children who had no access to e-learning infrastructure without options to continue their education. In this paper we present evidence from a large-scale para-teacher intervention which brought learning resources to the homes of children cut-off by school closures. Over the 6.5 months of intervention, children enrolled in the intervention saw an average increase in test scores of 1.87 SD, with greater gains for those with lower baseline assessment scores. With these gains achieved at a cost of 5.48-7.39 USD per SD, the intervention was extremely cost effective.*

KEYWORDS: primary education; Covid-19; lockdown; teachers; phygital learning

1. Introduction

The global shut-down in response to Covid-19 in March 2020 created a host of challenges for the education sector. As schools and educational institutions pivoted to remote learning, opportunities were discovered and weaknesses identified in areas which were previously off the radar. While online learning facilitated this difficult adjustment for many, those without access to e-learning infrastructure were left with few options. First generation learners, children from low socioeconomic backgrounds, children attending under-funded schools, and those living in low-income settings were particularly disenfranchised by the move to online or remote learning.

This paper presents evidence on an intervention that brought learning to the doorstep of school children caught on the wrong side of the digital divide when schools shut down. Developed as a stopgap solution by the Indian education NGO ASPIRE in collaboration with Tata Steel Foundation, the ‘Lockdown Learning’ intervention was originally designed to aid students they were already assisting before the pandemic. In January 2021, following nine

Correspondence Address: Margaret Leighton, Department of Economics, University of St Andrews, Castlecliffe, The Scores, St Andrews, Fife, KY16 9AR, United Kingdom. Email: mal22@st-andrews.ac.uk

Supplementary Materials are available for this article which can be accessed via the online version of this journal available at <https://doi.org/10.1080/00220388.2024.2337381>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

months of persistent school closures, the intervention underwent expansion to include an additional 100,500 children in grades 3 to 8, earning the moniker ‘Lockdown Learning Expansion.’

The intervention sent volunteer teachers, who were recruited from the local area, to visit participating students at home or in small groups. Each week, the teacher would bring a task and any necessary pedagogical supports to the children. The tasks drew on core curricular skills, such as math, writing and science, but were not part of the standard school curriculum: rather, they were hands-on assignments which required some degree of entrepreneurship. Students then worked independently, using locally available resources and support from their parents, neighbors, and peers to complete tasks. At the end of the week, the teacher returned to discuss, collect, and review the work.

In this study we estimate the learning gains that students made over the course of the Lockdown Learning Expansion. All participants completed a baseline learning assessment in January 2021. The intervention ran until the end of September 2021, with a 2.5 months stoppage from mid April to June end (the Delta wave of Covid-19 in India), for a total duration of 6.5 months. A 10.5% random sample of students was re-surveyed in the first week of October 2021 for an endline assessment.

Between baseline and endline, average scores on the assessment papers increased by 1.87 SD. These learning gains cost somewhere in the range of 5.48-7.39 USD per SD (alternatively, 13.53-18.25 SD per 100 USD), making the program extremely cost-effective. Improvements were slightly larger for upper primary students (grades 6-8) than primary (grades 3-5): 2.03 SD vs 1.75 SD respectively. We do not find any difference in learning gains across boys and girls, but do find evidence that students with lower baseline assessment scores showed larger absolute improvements over the course of the program than those with higher initial achievement.

These changes are estimated using a before-after study design.¹ With schools closed for the duration of the intervention, we argue that comparing the estimated learning gains from the intervention to a counterfactual of no change is conservative, with the true counterfactual likely to be a continued decline in academic skills over the 6.5 program months. Pre-pandemic evidence from school holidays and absenteeism indicates that learning declines substantially when schools are closed (see Kuhfeld et al. (2020)). More recently, Engzell et al. (2021) show that Dutch students made little or no progress when learning at home during Covid-19 school closures, with students from less-educated homes disproportionately affected. In Belgium, Maldonado and Witte (2022) find that schools with a higher proportion of disadvantaged students recorded larger pandemic-era learning losses. This suggests that even children from technologically advanced countries were struggling to learn remotely, with children from low resource families bearing the brunt of the loss.

While there is no doubt that learning has suffered during school closures, evidence on the extent and incidence of learning slow-down or learning loss - particularly in low income settings - is just starting to emerge. In the absence of real-time data, early estimates were generated using projections (see Angrist et al. (2021) in Sub-Saharan Africa, Khan and Ahmed (2021) in Pakistan, and Kuhfeld et al. (2020) in the United States). Bakhla et al. (2021) fielded a household survey in August 2021 in 15 Indian states (including Odisha and Jharkhand), focusing on underprivileged villages. Their findings suggest that nearly 50% of sampled children in rural areas were not studying at all at the time of the survey, while just 8% were studying regularly via online learning. Learning access rates were particularly bad for children from Scheduled Castes and Scheduled Tribes - marginalised groups in India. Learning levels were also very low among the sampled children. Bakhla et al. (2021) administered a simple literacy test to children and found that only half of children in the age group of 8-12 years could read a simple sentence. Only 25% of grade 3 children could read more than a few words.

One of the most accessible ways to reach children during Covid-19 school closures was by phone. Rigorous evaluations of phone-based pandemic interventions show mixed results.

Engaging parents with SMS messages or phone calls improved learning outcomes in Botswana by 0.12 SD (Angrist et al. (2022)); active phone calls by teachers or NGO staff lead to 0.14-0.19 SD improvements in learning in Nepal (Radhakrishnan et al. (2021)); while a 13-week tele-mentoring program improved learning outcomes in rural Bangladesh by 0.75 SD (Hassan et al. (2021)). In contrast, weekly tutoring phone calls had no impact on learning outcomes in Sierra Leone (Crawford et al. (2022)) nor, in a different design, in Kenya (Schueler and Rodriguez-Segura (2021)).

The intervention we evaluate here was implemented in a hybrid (online + physical) mode to maintain the continuity of learning for children who had no access to online learning. It is qualitatively different from the above studies as the volunteer teachers visited students at home, supported them in accessing and understanding digital resources, and gave them assignments independent of the school curriculum. Assigned tasks were geared towards transforming a child into a self-directed learner. We are not aware of any evaluations of interventions of this type supporting students during school closures, either in pilot studies or at scale. While our study is not based on a randomized control trial, it contributes to this literature by providing quantitative evidence on an intervention approach which shows great promise in a low-income, low-digital-access setting.

Our study makes three contributions. First, we report evidence on the effectiveness of a low-cost, scalable intervention which can support learning during times of institutional disruption. With conflict, environmental, and health risks ever-present, building an evidence base around such programs is important for resiliency. Second, the intervention we evaluate deviates from much of the Covid-19 impact evaluation literature in its reliance on face-to-face contact between teachers and students. While this is a limitation to some extent – the intervention was forced to pause during the most intense periods of the pandemic – we feel it has also been a strength. Particularly in low-income settings, but not exclusively, there are groups of children who need additional support to thrive as learners. Whether due to limited parental education or involvement, a disrupted educational journey for the child, or other contextual or personal factors, not all children can effectively learn remotely. The intervention we study offers an approach that is particularly suited to engaging children who are at a greater risk of falling behind.

Finally, this paper offers insights applicable to educational design beyond emergency situations. Many education systems are now considering the integration of digital learning solutions into the standard curriculum. In India, the recently introduced National Education Policy 2020 places particular emphasis on the “extensive use of technology in teaching and learning” as a fundamental principle for the future of the country’s education system (see Government of India (2020), page 5). Subsequently, the Ministry of Education launched PM eVidya, a comprehensive initiative to facilitate access to digital, online, and on-air education for all children. While these digital solutions hold great potential, the widespread push for their adoption raises concerns about the possibility of deepening the digital divide that has left many children behind during the pandemic. ‘Phygital’ learning, a blend of digital technology with face-to-face interactions, could serve as an inclusive approach, providing an on-ramp for children from diverse backgrounds. This intervention, as an example of phygital learning at scale, establishes an important precedent.

The remainder of the paper proceeds as follows. [Sections 2 and 3](#) present the context and describe the intervention. [Section 4](#) introduces the data, while [Section 5](#) provides an overview of the empirical methodology. Results, including main findings and heterogeneity analysis, are in [Section 6](#). [Section 7](#) provides some further discussion and cost-effectiveness calculations, and [Section 8](#) concludes.

2. Context

In 2020, in response to the Covid-19 pandemic, governments across the world closed schools and educational institutions, adversely affecting 1.6 billion children and young people

(UNESCO (2021)). While India followed the global trend by closing schools early in the pandemic, it soon stood apart, implementing one of the longest continuous school closures in the world at 69 weeks on average (The Economist (2021)).

In the states of Odisha and Jharkhand, schools across all classes closed a week prior to the nationwide lockdown imposed on March 24, 2020. The national lockdown gradually relaxed over the next few months as decisions regarding Covid-19 containment measures were handed to states. Two successive waves of the virus swept India over the subsequent 18 months—the Delta and Omicron variants—with the Delta wave proving particularly lethal in many states. Consequently, containment policies underwent abrupt changes, and school closures were often among the initial measures enacted by state governments.

In February 2021, both Odisha and Jharkhand cautiously reopened for grades 9 to 12, anticipating subsequent re-openings for additional grades. However, the Delta wave struck in March-April of that year, prompting the closure of schools for all grades. Another phased reopening for grades 6 to 12 began in October 2021. Three months later, the Omicron variant caused an alarming increase in cases. In response, both state governments once again mandated the closure of schools from January 2022. Finally, schools reopened for all grades for the first time from July 2022. From March 2020 to June 2022 government schools in Odisha and Jharkhand remained physically closed for primary students (grades 1 to 5) on all days.

To maintain the continuity of learning, Indian state education departments responded with various distance learning solutions: e-learning platforms, video classes, messaging apps, learning apps, television programs, and public radio programs. Public schools in Odisha and Jharkhand adopted e-learning apps such as Madhu and DIKSHA, utilized community radio stations, and engaged with e-learning platforms like e-Pathsala and e-Mulyankan to deliver educational content to students. Exploring time-use data from a nation-wide panel, Andrew and Salisbury (2023) find that Indian children age 12-18 continued to study during school closures, albeit at much lower rates than pre-pandemic, suggesting that these measures had some limited success at reaching older students.

However, not all learners and teachers were ready for this transition. While grades 9-12 experienced intermittent school openings that provided older students with some access to learning in classrooms, this was not the case for grades 3 to 8. Public schools for these grades remained physically closed for the vast majority of days between March 2020 and June 2022. The absence of proper digital infrastructure, digital literacy, internet-enabled devices, and conducive home environments presented unprecedented challenges for most of these young students and their instructors.

Recognizing the limited digital access, public school teachers in Odisha and Jharkhand were directed to distribute textbooks to grade 1-8 children at their doorsteps for self-study. While this was well-intended, it did not translate into effective engagement for the majority of students. A survey carried out by ASPIRE late in 2021 found that, of enrolled children age 7-14, only 8.8% received learning materials or activities from their school teacher at least once a week, while 51.2% never received anything. Public school students, who make up 85.3%, fared even worse, with only 5.2% receiving materials weekly.

3. Intervention

3.1. Overview and background

ASPIRE, an Indian NGO, operates in remote and underdeveloped indigenous and tribal districts of Central and Eastern India (Odisha, Jharkhand, Chhattisgarh, and West Bengal states). The organization aims to revitalize the public education system, with a particular emphasis on preparing children for the challenges of the 21st century. Since 2015, these efforts have come under the umbrella of the NGO's wide-ranging Education Signature Program.² This program combines multiple interventions implemented at scale, simultaneously addressing three aspects

of schooling: access, learning, and governance. The learning arm incorporates a range of activities, including a learning enrichment program designed to enhance foundational literacy, training government teachers in effective pedagogy, conducting remedial classes to address learning gaps in math and science, promoting the use of digital technology and computational thinking, and establishing school libraries.

As schools physically closed in late March 2020, ASPIRE initiated a village-level digital access mapping to identify students with internet access or household smartphone availability. Focused on the 30,000 children enrolled in ASPIRE's learning enrichment program at the onset of the pandemic, the mapping revealed that only 14% of these children had either smartphone or internet access at home. Recognising this education emergency, ASPIRE, with funding from Tata Steel Foundation, launched a hybrid (online + physical) learning intervention in its program areas to maintain continuity of learning.³

While this Lockdown Learning (LL) program targeted ASPIRE's existing students, siblings of these children of a similar age were also invited, raising the total enrollment to approximately 35,000. The existing curriculum-based teaching-learning material from the learning enrichment program was redesigned to engage a child at home, on her own, with support from parents, neighbors, and peers, and using resources available locally. 700 ASPIRE teachers were each assigned a set of villages, where they would meet children twice a week, in an open air environment, individually or in small groups, following mask, sanitization, and distancing protocols. On the first visit of the week, teachers introduced the learning task on their phones or tablets and on the second visit, completed tasks were reviewed and collected.

In January 2021, with schools still closed, the LL program was supplemented by a Lockdown Learning Expansion (LLX). Across all project areas, an invitation was extended to all students in grades 3 to 8 without access to learning. This excluded those already enrolled in the LL program, receiving private tuition, or having regular access to e-learning from their schools.⁴ 100,500 children joined in this new program, approximately 34% of all children in grades 3-8. These children were spread over 5,555 villages in 19 blocks in Odisha and Jharkhand. A team of 2,700 volunteer teachers was temporarily recruited and trained to support and augment the existing 700 ASPIRE teachers.⁵

After launching in January 2021, the LLX experienced a 2.5-month hiatus from mid-April to the end of June 2021 due to the Delta Covid-19 wave. Visits then resumed until the end of September 2021 when the intervention concluded, coinciding with the states' planned reopening of schools for all grades from October 2021.⁶

3.2. Implementation structure

Students across the program worked on the same sets of tasks at the same time, adapted for age. At the start of each week, on Sunday, volunteer teachers received orientation for the week's task from their respective supervisors. On Monday and Tuesday, they visited their assigned villages and met the enrolled children. They introduced the learning tasks to the children using their phones or tablets. On Wednesdays and Thursdays, if needed, volunteer teachers revisited their assigned villages to assist students who had struggled to complete tasks in the preceding week. Fridays and Saturdays were designated for collecting completed tasks. On average, a volunteer teacher spent 15-20 minutes per child and reached out to 30-40 children across 2-4 villages. The average class size was 5-7 children. In some areas, children were approached individually, where in a few areas, groups of 40-50 students were created. All classes were held outdoors. Volunteer teachers were trained to follow certain protocols during their visits: address each child by name, try to know their background, maintain a good relationship with the parents, and encourage a respectful, fear-free interaction.

A multi-tiered implementation and monitoring structure was set up to run the LLX, ensuring all students received the intervention to the same standard. The 700 ASPIRE teachers

supervised and supported the volunteers, offering ongoing feedback and preparing tasks for the following week's cycle. For every 5 ASPIRE teachers, a coordinator was assigned, responsible for monitoring, data collection, and providing pedagogical guidance. Each of the 19 blocks had a designated manager overseeing the intervention logistics. Ten percent of the completed tasks were uploaded and shared with senior management at ASPIRE and Tata Steel Foundation for review and feedback.

The LLX started with simple, subject-specific learning tasks and advanced to small projects that integrated various subjects and skills. An early hybrid task, for instance, involved conducting an internet search on a specific topic like giraffes, Mount Everest, dinosaurs, or the Taj Mahal and creating a drawing, write-up, or presentation by the end of the week. For students with no smartphone or internet access at home, the volunteer teacher would allow them to use her smartphone during the initial visit. In areas without internet access, volunteer teachers would use pre-downloaded lessons and videos to explain the task.

A great diversity of learning tasks were designed and sent to the students. Children browsed the Johns Hopkins University coronavirus resource page to discuss the math and science of Covid-19, practiced math concepts at the neighborhood shop, observed leaves and insects in their surroundings to understand biodiversity, and listed tasks carried out by their mothers for gender equality discussions. In a citizenship task, grade 6-8 children spread awareness about Covid-19 vaccines and helped register 16,000 local people for vaccination. The total time spent by children on tasks depended on the type of task. Some tasks were simple search-and-write tasks, while others involved venturing into the local environment to observe and collect information. Excluding the preparation time, a child spent roughly 20-30 minutes on the final write-up for a simple task and 50-60 minutes for a higher-level task. A detailed example of three tasks is given in the Supplementary Material.

The intervention also promoted parental and community involvement in their child's learning through tasks such as compiling kitchen recipes, documenting stories from grandparents, and learning village history from elders. In 70%-80% of villages, children organized exhibitions to showcase their work to the community, and parents were invited to discuss and reflect. This was novel for these communities as the majority of the children are first-generation learners, and previously their parents were not actively engaged in their education.

4. Data

4.1. Data collection

A baseline assessment was conducted for all 100,500 children who enrolled in the LLX program. Tests were administered in the local language - Odia for Odisha and Hindi in Jharkhand. These baseline assessments also collected gender, grade level at school, and social category status. A 10.5% random sample of these children were selected for follow-up at endline, resulting a panel of 5192 primary and 4275 upper primary grade children.

The baseline and endline assessments used the same papers. Primary grade children received two papers: a math assessment aligned with the government curriculum for grade 3 and a language assessment benchmarked to grade 5. These primary grade papers are akin to those used by ASER for nationwide learning level monitoring. The upper primary grade assessment, tailored for self-directed tasks, evaluated various skills across eight topics.⁷ Guidelines for grading baseline and endline papers were developed and communicated to volunteer teachers. Papers were graded by volunteers, with a random sample cross-checked by supervisors and re-evaluated by block managers. Data entry in Excel was conducted by monitoring staff.

The decision to use the same paper raises some concerns with respect to identification. It is possible that changes in assessment scores could be inflated by baseline exposure to the assessment paper, either by the students (e.g. through memorisation of questions) or by the teachers (e.g. by focusing on an overly narrow set of material). While we cannot rule out these risks,

four features convince us that the magnitude of such an effect will be modest. First, a full 9 months elapsed between baseline and endline, which is a long time for a child to remember test questions. Second, the assessment papers were not corrected and returned to the children to study. Third, these were low-stakes assessments which had no bearing on program eligibility or future participation. Fourth, the teachers themselves received no remuneration based on assessment paper scores.⁸

4.2. Overview of the sample

While our sample is a random selection of LLX participants, children enrolled in the program are not a representative sample of all children in their communities. Eligibility for the program was determined based on need and the child's history with existing ASPIRE interventions. Children previously enrolled in the learning enrichment program were already receiving an early version of the LLX, and were therefore excluded: these were among the worst academic performers at their schools. Children who had reliable access to learning, either through private tuition, regular internet access at home, or other means, were also excluded from the LLX. These children are likely to include the most academically advantaged children. Further details on the sample and attrition are provided in the Supplementary Material (Section 1.2).

An overview of the sample characteristics is presented in Table 1. Girls and boys are equally represented in both the primary and upper primary samples. Due to the location of the intervention in tribal areas, the sample is dominated by children from Scheduled Tribes (ST): 67% of the primary sample and 63% of the upper primary sample come from this population. Other Backwards Castes (OBC) make up 22% of the total and Scheduled Castes (SC) make up 10%. Children from general castes make up only 3%.

The intervention offered two curricula: one for primary grades 3-5, and the other for upper primary grades 6-8. The vast majority of students in the program complied with these cut-offs, although there were a few exceptions, e.g. grade 6 children following the primary course. Students enrolled in the primary course are quite evenly spread across grades 3-5, with only slightly more enrolled in 3rd grade than in other grades. The upper primary group is more heavily skewed towards the younger grades: 39% of the sample are in grade 6, compared with only 27% in grade 8.

Table 1. Summary statistics: sample

	Total		Primary		Upper Primary	
	Mean	SD	Mean	SD	Mean	SD
Girls	0.50	0.50	0.50	0.50	0.50	0.50
General	0.03	0.16	0.02	0.15	0.03	0.17
OBC	0.22	0.41	0.20	0.40	0.24	0.43
SC	0.10	0.30	0.10	0.30	0.10	0.31
ST	0.65	0.48	0.67	0.47	0.63	0.48
Grade 3	0.20	0.40	0.36	0.48	0.00	0.00
Grade 4	0.17	0.38	0.31	0.46	0.00	0.02
Grade 5	0.18	0.38	0.32	0.47	0.00	0.03
Grade 6	0.18	0.38	0.01	0.07	0.39	0.49
Grade 7	0.15	0.36	0.00	0.00	0.34	0.47
Grade 8	0.12	0.33	0.00	0.00	0.27	0.44
Grade 9	0.00	0.03	0.00	0.00	0.00	0.04
N	9,467		5,192		4,275	

Abbreviations: General = general caste; OBC = Other Backwards Caste; SC = Scheduled Caste; ST = Scheduled Tribe. This table is discussed in Section 4.2.

To address the non-representative nature of the LLX cohort, we seek to put them into context within their communities by drawing on a household survey carried out by ASPIRE in a subset of our study area. (Details of the household survey, which was carried out in late 2021, can be found in the Supplementary Material, Section 1.2.3.) A comparison within their communities shows that LLX children come from less educated and less well-equipped households (SM Table 11): their parents have completed fewer years of education (fathers: 4.9 vs 6.0 years; mothers 3.8 vs 5.1 years), and they are less likely to have access to the internet (0.28% vs 0.40%) or a smartphone (0.17% vs 0.26%) at home. LLX children are more likely to be enrolled in public schools (97% vs 80%) and come from marginally larger families (5.25 vs 5.17). These statistics suggest that, while on average LLX children come from backgrounds with weaker support structures for their education, they remain broadly comparable to their peers who did not attend the LLX.

4.3. Outcome variables

Our primary outcome of interest is the overall assessment score. This score is the average of math and language papers for primary school children, and a weighted average of the eight competencies for upper primary children.⁹ Table 2 provides an overview of the raw scores at baseline and endline. The overall total score at baseline for primary school students was 27%, the average of 31% for math and 24% for language. Upper primary students had slightly lower scores at baseline, with an overall average of 20%. Upper primary students scored an average of 21% in math and 15% in writing, with particularly high scores on the shapes and materials topic (41%), and lowest average scores on Covid-19 knowledge (11%).

Endline scores are considerably higher for all groups, in all subjects. Primary school students increased their mean score by 44 points to achieve an average of 71%; upper primary students increased theirs by 40 points for an endline average of 60%. Increases are similar across math and language for primary students. There is more variation in subject-level increases in upper primary, from a 26 point increase in Covid-19 knowledge to a 59 point increase in maps. For core subjects, upper primary students saw an increase of 32 points in writing, and 48 points in math.

For analysis, we normalise the assessment score data with respect to the baseline in each group. For our headline results, we use normalised total score as our outcome variable, and aggregate primary school and upper primary students together. In other analyses, we split the two groups, and also investigate the subjects individually.

Table 2. Summary statistics: assessment scores

	Primary School				Upper Primary School			
	Baseline		Endline		Baseline		Endline	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total	26.99	25.34	71.26	20.96	19.98	19.72	60.00	20.82
Math	30.65	30.93	78.43	23.66	20.54	31.06	68.17	35.36
Language	24.09	24.37	64.09	22.61				
Writing					14.89	21.57	46.81	26.91
English					18.64	22.78	54.18	27.55
Science					14.87	24.22	55.22	32.14
Shapes Materials					41.27	32.82	82.40	20.42
Calendar					23.54	35.88	74.42	36.55
Maps					21.20	28.50	81.00	26.20
Covid					11.28	20.66	36.95	28.42

Note: This table is discussed in Section 4.3.

5. Empirical approach

Our analysis is based on a pre-post design. Given the pandemic conditions under which the intervention was developed and implemented, it was not possible to collect data on an untreated control group while the intervention was running. Without a control group, we do not have a precise counterfactual of how assessment scores would have evolved over this time period, in the absence of the intervention.

What we do know is that, for the duration of the intervention, government schools were closed for grades 3 - 8. Furthermore, children selected for the intervention were those who had no other learning opportunities available to them. To complement our estimated average improvements in test scores over the course of the intervention, we extend our analysis with a discussion of possible counterfactuals, drawn from the recent literature on learning loss during the pandemic (Section 7.1).

Our primary estimating equation is as follows:

$$y_{it} = \beta_0 + \beta_1 Post_{it} + \Gamma X_i + \epsilon_{it}, \quad (1)$$

where y_{it} is the assessment score of student i at time $t = 1, 2$; $Post_{it}$ is a dummy variable equal to 1 in the post period; and X_i are a set of time-invariant individual characteristics: sex, school grade at baseline, and social category (General, OBC, SC or ST). ϵ_{it} is the unobserved error term. Our coefficient of interest is $\hat{\beta}_1$, the empirical estimate of the difference in baseline and endline assessment scores.

6. Results

6.1. Main results

Our primary outcome of interest is the normalised total assessment score of each student. Table 3 presents estimates of Equation 1 with the total score as the outcome variable, measured in standard deviations of the baseline test score. Column 1 combines primary and upper

Table 3. Main results: total score in standard deviations

	(1)		(2)		(3)	
	Full Sample		Primary		Upper Primary	
Post	1.874***	(0.0800)	1.747***	(0.0835)	2.029***	(0.0998)
Girls	0.0636***	(0.0196)	0.0663**	(0.0236)	0.0590**	(0.0273)
OBC	-0.186*	(0.0985)	-0.113	(0.107)	-0.260**	(0.117)
SC	-0.426***	(0.114)	-0.432***	(0.114)	-0.410***	(0.137)
ST	-0.499***	(0.112)	-0.481***	(0.113)	-0.514***	(0.129)
Grade 4	0.181***	(0.0184)	0.179***	(0.0180)	1.516***	(0.102)
Grade 5	0.293***	(0.0290)	0.294***	(0.0285)	-0.202	(0.195)
Grade 6	0.182***	(0.0542)	-0.0699	(0.0616)		
Grade 7	0.322***	(0.0432)			0.137***	(0.0267)
Grade 8	0.349***	(0.0514)			0.163***	(0.0398)
Grade 9	-0.109	(0.113)			-0.291***	(0.0979)
Constant	0.173	(0.109)	0.209*	(0.101)	1.824***	(0.151)
Observations	18925		10375		8550	

Notes: each column reports a separate OLS regression estimating Equation 1 with total score (in standard deviations of the baseline score) as the outcome variable. Column 1 includes the full sample, Column 2 primary only, Column 3 upper primary only. Omitted categories: Pre-period, Boys, General caste, youngest expected grade (grade 3 for Columns 1 and 2; grade 6 for Column 3). Standard errors in parentheses clustered at the block level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table is discussed in Section 6.1.

primary students, while Columns 2 & 3 consider each of them individually. Overall, assessment scores increased by 1.87 SD between baseline and endline. Primary school students had slightly smaller relative gains, at 1.75 SD, compared to upper primary students with a 2.03 SD improvement.

Keeping in mind that these regressions pool pre and post data,¹⁰ some of the differences by student characteristic are worth noting. On average, girls perform slightly better than boys (by 0.06-0.07 SD). Compared with the general population (which is a minority in many parts of the study region), OBC, SC, and ST students perform worse on average, with SC and ST most substantially behind by 0.4-0.5 SD.

The differences in performance by school grade level are surprisingly modest. As the data are normalized within assessment papers (primary and upper primary separately), we limit ourselves to comparison within these grade bands. Primary school students have an average difference of 0.15 SD per grade level, while the scores of upper primary school students are more compressed: 7th graders score 0.14 SD higher than 6th graders, but 8th graders score only 0.03 SD higher than 7th graders. It could be that the year-to-year learning in remote communities such as these is very small, or that after a full year of school closures the upper primary students had lost more than others. It could also be that high-performing 8th graders were more likely to be taking private tuition, and therefore not enrolled in the LLX.

Table 4 presents estimates of the treatment effect on each subject individually. With the exception of Maps, the subject-specific relative increases are smaller than the overall score increases shown in Table 3. These lower standardised increases can be explained through a combination of greater variation in the subject scores than the total score, and smaller increases in assessment scores in some subjects than others. For example, the standard deviation in math scores on the primary school assessment was 30.9 points, compared to a standard deviation of 25.3 for the total score (see Table 2): a similar percentage point increase in math scores will be normalised into a smaller standardised increase. As shown in Table 2, the percentage point increases also vary considerably.

Although the effect sizes are smaller than for the overall score, the gains in individual subjects remain substantial: in all cases these are greater than 1 SD, with most around 1.5 SD, including the core subjects of math, language, and writing. The smallest relative gains are seen in the Shapes and Materials subject (Column 7) and Covid knowledge (Column 10).

6.2. *Robustness*

We estimate three alternative specifications. First, our main results are re-estimated with child fixed-effects. These estimates are almost identical to the primary specification (SM Table 3). Next, we replicate our analysis using an aggregate of math and languages scores only: while the overall estimate drops marginally to 1.76 SD, the results are highly similar (SM Table 4). Finally, we check that our results are not driven by normalisation by re-estimating our overall and subject specific effects using percentage point scores (SM Tables 5 and 6). The average change in score across the sample is 42 percentage points.

6.3. *Heterogeneity*

6.3.1. *Student characteristics.* Which students benefited the most from the program? To estimate heterogeneity in our treatment effects, we re-estimate Equation 1 fully interacted with selected student characteristics.¹¹ We focus on three characteristics: child gender, child grade, and baseline assessment score.

Results are presented in Table 5. The first row of each column presents the estimated learning gains for the omitted category, with subsequent rows showing estimated differences with respect to that category. The results indicate that there are no significant differences in learning gains

Table 4. Subject results: score in standard deviations

	Upper Primary									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Math	Language	Math	Writing	English	Science	Materials	Calendar	Maps	Covid
Post	1.545*** (0.0785)	1.641*** (0.0885)	1.534*** (0.0916)	1.480*** (0.0920)	1.560*** (0.0865)	1.666*** (0.102)	1.253*** (0.0606)	1.418*** (0.0821)	2.098*** (0.0823)	1.243*** (0.131)
Girls	0.0280 (0.0170)	0.104** (0.0302)	0.00738 (0.0240)	0.126*** (0.0306)	0.0210 (0.0268)	0.0277 (0.0336)	0.0279 (0.0163)	0.0412 (0.0281)	0.00271 (0.0227)	0.0892*** (0.0278)
OBC	-0.102 (0.111)	-0.179 (0.137)	-0.217 (0.110)	-0.253* (0.103)	-0.155 (0.0996)	-0.208 (0.128)	-0.115 (0.0898)	-0.227* (0.0821)	-0.216 (0.113)	-0.163 (0.119)
SC	-0.371** (0.115)	-0.481*** (0.120)	-0.372** (0.129)	-0.345** (0.111)	-0.336** (0.109)	-0.322* (0.139)	-0.177 (0.109)	-0.347** (0.121)	-0.269 (0.129)	-0.253* (0.105)
ST	-0.421** (0.120)	-0.558** (0.153)	-0.437** (0.115)	-0.486*** (0.118)	-0.482*** (0.105)	-0.311 (0.151)	-0.313** (0.104)	-0.265** (0.0833)	-0.385** (0.126)	-0.319* (0.125)
Observations	10376	10376	8550	8550	8550	8550	8549	8550	8550	8537

Notes: each column reports a separate OLS regression estimating Equation 1 with subject score (in standard deviations of the baseline score) as the outcome variable. Columns 1-2 are Primary only, Columns 3-10 are Upper Primary only. Omitted categories: Pre-period, Boys, General caste; other controls (not shown): indicator for grade at baseline. Standard errors in parentheses clustered at the block level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table is discussed in Section 6.1.

by gender (Column 1); there are, however, differences by grade (Columns 2 & 3). For both primary grades and upper primary grades, the largest gains are seen by the youngest grade level. A similar trend is seen with baseline assessment scores (Column 4): those with higher initial performance gained less. Specifically, a 1 SD higher baseline assessment score is associated with 0.70 SD smaller gains in test scores between baseline and endline.

It could be that the older children, and those with higher baseline scores, had less potential to benefit from the intervention: or conversely, those with very low initial learning levels had more room to improve. A similar trend might emerge if the curricula were pitched towards the lower-achieving students in each group. It could also be that there is negative selection into the program among the older and higher-achieving students based on characteristics associated with academic success: for example, families who put a high value on education might make particular sacrifices to help older children access private tuition during the pandemic.

6.3.2. Community characteristics. Does the effect of the program vary based on characteristics of the communities? To explore whether there are meaningful differences in learning gains by community characteristics we draw on data from the 2011 Indian census covering socioeconomic characteristics of households as well as education infrastructure. Although the census has such data at a fine level of geographic detail, we are only able to merge this with our analysis sample at the block level. Restricting to blocks within Odisha (92.3% of our sample), our study sample spans 17 blocks.¹² Summary statistics are provided in the Supplementary Material (see Table 7).

We explore the impact of block-level characteristics on our estimates by re-estimating our main results with block-level controls, and also by re-estimating Equation 1 fully interacted with selected block characteristics. In the first case, when controlling for block-level population, area, literacy rate and number of primary schools, we find no difference in our main results (see SM Table 8). The interacted models (male and female literacy rates, number of schools) are shown in Table 6. While none of the interaction effects we estimate is statistically different from

Table 5. Heterogeneity by student characteristics: total score

	(1) Gender	(2) Primary Grade	(3) Upper Primary Grade	(4) Baseline score
Post = 1	1.890*** (0.0776)	1.865*** (0.0911)	2.082*** (0.105)	1.874*** (0.0864)
Post = 1 × female	-0.0310 (0.0233)			
Post = 1 × Grade = 4		-0.103** (0.0338)		
Post = 1 × Grade = 5		-0.256*** (0.0375)		
Post = 1 × Grade = 7			-0.0424 (0.0486)	
Post = 1 × Grade = 8			-0.135** (0.0427)	
Post = 1 × Baseline score				-0.669*** (0.0398)
Observations	18925	10323	8528	18924

Notes: each column reports a separate OLS regression estimating Equation 1 with total score (in standard deviations of the baseline score) as the outcome variable. Each regression includes the following controls, fully interacted with the heterogeneity variable: gender, social category and grade. Heterogeneity variables are: gender (Column 1), grade (Columns 2 & 3) and baseline assessment score. Post = 1 represents the learning gains for the omitted category: male (Column 1), grade 3 (Column 2), grade 6 (Column 3). In Column 4 the heterogeneity variable is continuous. Standard errors in parentheses clustered at the block level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table is discussed in Section 6.3.

Table 6. Heterogeneity by community characteristics: total score

	(1) Female literacy	(2) Male literacy	(3) Primary schools
Post = 1	1.630*** (0.376)	1.413* (0.586)	2.110*** (0.343)
Post = 1 × Block female literacy rate	0.539 (0.827)		
Post = 1 × Block male literacy rate		0.736 (0.940)	
Post = 1 × Block primary schools			-0.00148 (0.00221)
Observations	17249	17249	17249

Notes: each column reports a separate OLS regression estimating Equation 1 with total score (in standard deviations of the baseline score) as the outcome variable. Each regression includes individuals controls (gender, social category, and grade) and basic block controls (total area, total population, population aged 0-6). Standard errors in parentheses clustered at the block level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table is discussed in Section 6.3.

zero, they are suggestive about the influence of community-level factors. In particular, the interaction with male and female literacy rates suggest a positive association between adult literacy and learning gains over the course of the intervention. The interaction with number of primary schools is small and insignificantly negative.

7. Discussion

7.1. Plausible counterfactuals

What happens to learning when children are out of school? What is an appropriate counterfactual for the children we observe before and after the intervention? In a review of pre-pandemic studies of learning loss, Kuhfeld et al. (2020) find daily losses from summer holidays or absenteeism for grades 3-5 in the range of 0.005-0.007 SD for math and 0.003-0.005 SD for English; for grades 6-8, 0.002-0.008 SD for math and 0.001-0.006 SD for English (Kuhfeld et al. (2020): Table 1).

Starting in January 2021 and running until the end of September, with a break from mid April-June, the LLX ran for 6.5 months, or approximately 200 days. If the program had not run, children would have been fully out of school during that time. If we apply the estimates of summer or absentee learning loss from Kuhfeld et al. (2020), we would expect a loss for primary students of 1-1.4 SD in math and 0.6-1 SD in language, while for upper primary the equivalents would be 0.4-1.6 SD for math and 0.2-1.2 SD for language.

It is likely that these extrapolations overstate the learning loss the LLX students would have experienced over the period of the intervention. At baseline, these students had been out of school for nine months, and may already have lost nearly all of the math and literacy skills they had learned at school. This seems to have been the case for part of our sample: 10.3% of primary students and 14.2% of upper primary students score a 0 on the baseline assessment. These students had no remaining learning to lose by continued school closures.

We expect therefore that the true counterfactual evolution of learning for the students in our sample, in the absence of the intervention, lies somewhere between 0 and -1.6. If we take the example of primary school language, our point estimate is an increase of 1.64 SD (Table 3). Our extrapolation from Kuhfeld et al. (2020) suggest a loss of 0.6-1 SD in language skills for this age group over 6.5 months. If that is the true counterfactual for our students, then the causal effect of the intervention would be in the range of 2.24-2.64 SD. If, however, the true

counterfactual is a learning loss of 0 the causal effect of the intervention is our point estimate itself, +1.64 SD.

Could students have made learning gains over this period, in the absence of the intervention? With schools closed, and no access to private tuition or other learning options, it is unlikely that many of the children in the sample would have achieved positive learning gains over the period. There are no doubt some exceptions in the 100,500 children enrolled in the intervention: however, the vast majority would have seen a continued decay in their academic skills while they remained cut-off from education.

7.2. *Contextualising the results*

Our results showcase the impact of a unique learning intervention implemented during Covid-19 school closures. The intervention targeted children with limited access to internet and alternative avenues of learning. While a 1.87 SD increase in test scores is large compared to educational interventions studied prior to the pandemic, such magnitudes are not unheard of.¹³ However, when interpreting these results it is important to keep in mind that most education interventions evaluated prior to 2020 have been marginal to some existing school system. The intervention we study supported students in an institutional vacuum.

The effect sizes are nevertheless substantial. Several unique features of the intervention might be driving the results. Angrist et al. (2022) find that parental engagement in a child's education improves learning outcomes during school closures, even in the low literacy context of Botswana. As highlighted in Section 3, many of the intervention tasks invited participation of parents. Children shadowed their mothers all day at home and in the fields to document their activities for the task "what mothers do"; they went with their parents to the local weekly market for a photo-story task; and to the local grocer to learn math concepts. For many of the parents, who had not attended school themselves, this was their first time engaging with their child's education. Anecdotal evidence from the field suggests that parents were grateful for the learning support and felt their children were neglected by the government teachers. They voiced excitement over understanding first-hand their child's effort and learning level.

Children were also learning outside of the rigid structure of classrooms and textbooks, and at their own pace. Forests, ponds, farms, homes, and markets became learning spaces, where children practiced math and science, increased their awareness around gender, climate change, and biodiversity. Evidence from health and psychology points to the benefit of unstructured learning, especially outdoor play time, on the well-being of pre-school and young primary grade children (see Brussoni et al. (2017) and Lee et al. (2020)).

One of the objectives of the intervention was to motivate children to become self-directed learners. Tasks were designed to engage a child alone: they were exploratory in nature, and they invited children to use the internet and resources available locally. Evidence-based research on the benefits of self-directed learning is still nascent (see Brandt (2020) for a review). More research is needed to understand to what extent the large gains in student test scores from this intervention are a result of the self-directed learning process, and whether similar gains could be obtained by fostering self-directed learning in other contexts.

Our finding that low achievers had disproportionately larger learning gains over the course of the intervention is particularly relevant as more and more countries are leaning towards incorporating e-learning solutions into their education systems. Previous evidence has found that low achieving children compensated less for school closures than did their higher-achieving peers (Grewenig et al. (2021)), aggravating inequality in educational outcomes. Combining high quality digital learning materials with personalised, attentive teacher-student contact could be a way of reaping the rewards of technology-driven education initiatives without worsening educational inequality across the digital divide.

7.3. Spillovers

Children influence each other, and it is important to consider the interplay between those enrolled and not enrolled in an intervention. In our context, there could be spillovers from the LLX onto non-participant children; there could also be spillovers from non-participant children onto our treatment group. The first case does not threaten our identification of the effect of the program on enrolled children; the second case does.

Spillovers from non-participants onto LLX children could be positive or negative. Negative spillovers could occur if non-LLX children were particularly discouraged about their own education, perhaps as a result of the difficulty posed by remote learning even for those who had access to the internet. Such disengagement could spread to other children, and create a culture of disinterest in school-related work.

Positive spillovers, on the other hand, might occur due to non-LLX children having greater access to educational opportunities: if such children were continuing to learn and study despite school closures, they could be supporting, directly or indirectly, the learning of their peers who had more restricted access to remote education. If these effects are substantial, then our hypothetical counterfactual, wherein LLX children would not be learning in the absence of the intervention, could be overly pessimistic. In such a case, our estimated learning gains would be due to a combination of the LLX and spillovers from non-program peers.

Our data do not allow us to rule out such spillovers. What we can say empirically is that the academic level of LLX children at the start of the intervention was very low, and rose dramatically by the time the endline data collection took place. The very low scores on the baseline assessment indicate that, if there were positive spillovers from children with better access to remote learning onto the LLX children with little to no access, these did not translate into substantial learning gains during the first year of school closures. If any of the subsequent learning gains measured over the course of the intervention were due to positive spillovers from non-LLX children, we conclude that these were modest at most.

7.4. Cost effectiveness

Between January and September 2021, the combined budget for the original Lockdown Learning (LL) and LLX interventions amounted to 101,390,625 INR (1,389,052 USD as of January 2021).¹⁴ This budget encompassed various expenses, including honorariums for 2,700 volunteer teachers; salaries for 700 ASPIRE teachers and their supervisors; training costs; and teaching-learning materials. Throughout the six months of the LLX intervention, volunteer teachers were exclusively funded for the 100,500 LLX children. However, the 700 ASPIRE teachers and their supervisors divided their time between both LLX and LL, which ran concurrently and included a total of 35,000 children. Precisely budgeting the time allocated by the 700 ASPIRE teachers to the two groups presents a challenge.

Allocating the total budget exclusively to LLX children results in a per-child cost of 1009 INR (13.82 USD), while a shared allocation across both LL and LLX yields a per-child cost of 748 INR (10.25 USD). With a cost range of 10.25-13.82 USD per child and a primary estimate of 1.87 SD improvement per child, the cost-per-SD improvement is 5.48-7.39 USD (alternatively, 13.53-18.25 SD per 100 USD).¹⁵

In the context of pre-pandemic learning interventions, these cost effectiveness estimates are quite astonishing (see, e.g., Kremer et al. (2013)). There are two important points to keep in mind when interpreting these numbers. First, ASPIRE already had an established presence in the areas where the intervention was rolled out. The costs estimates are therefore at the margin, assuming an established presence. Second of all, this is a not a normal-times education intervention supplementing an existing education system: rather, it provided some amount of ongoing education during a time of institutional failure. Given the educational vacuum in which children found themselves at this time, the effectiveness of the intervention is not too surprising.

8. Conclusion

In a scramble for normalcy under Covid-19 lockdown, parents, teachers, civil society, and governments sought to ensure continuity in education in the face of unprecedented challenges. The variety of policy innovations which emerged in response to school closures provide a wealth of insights into ‘what works’ in education. Unfortunately, the very conditions which gave rise to these innovations make it difficult to rigorously evaluate them. As an education NGO with an established grass-roots presence, ASPIRE was uniquely positioned to design and implement a large scale intervention to support children during school closures. While the crisis conditions under which the intervention took place precluded the collection of data on a control group in a non-program area, monitoring data from the intervention help us paint a picture of how it affected children.

Our study estimates the learning gains achieved by children over the 6.5 month Lockdown Learning Expansion intervention. We apply a before-after analysis to panel data from over 17,000 students who, absent this program, would have had no access to education. We estimate an average improvement in test scores of 1.87 SD. While the gains are similar for boys and girls, they are larger for marginalised or vulnerable groups, including Scheduled Tribes and children with particularly low test scores at baseline. Using our headline estimate as the treatment effect, the intervention delivered improvements in learning at a cost of 5.48-7.39 USD per SD (13.53-18.25 SD per 100 USD).

This paper has a number of limitations. First, our reliance on a before-after design provides a weaker causal attribution of change than we would like. While we argue that, in this particular context, a counterfactual of zero change is actually fairly conservative, this is not the same as being able to follow a control group. Questions remain about the true counterfactual evolution of assessment scores.

Second, while the intervention was offered to all children who met basic eligibility criteria, it was not a universal program. In particular, children who were part of ASPIRE’s learning enrichment program prior to the pandemic (among the weakest students) were excluded, as were those who had access to private tuition or other education opportunities (likely to be among the better performers). Those who did participate were therefore not representative of the full population: our results cannot be interpreted as population mean effects.

Finally, the intervention we study is a complex one: it brought new resources and educational approaches, engaged parents and communities. By encouraging initiative and self-directed learning, and by showcasing the talents of students that might never have been noticed before, it also had the potential to build agency and academic self-esteem. While this complexity surely contributed to its effectiveness, it makes it difficult to piece out the most important channels of effect. Our data have little to say about which aspects of the intervention were critical to success. This also makes it difficult to extrapolate the successes measured here to other contexts.

ASPIRE’s highly cost-effective Lockdown Learning Expansion has a number of important lessons for policy. First, it offers a model of physical + digital learning which could be rolled out at scale during future periods of institutional disruption. It also offers insights into how the digital divide could be breached, supporting those students who are at risk of being left behind by the increasing reliance on digital learning technologies. This is a pressing issue for policy makers who face an equity-efficiency trade-off in their mission of improving education quality in low-income settings.

Our findings point to a number of directions for future research. First, focusing on this specific intervention, it would be useful to know more about the heterogeneity of improvements. While the monitoring data contains only minimal information on the individual students, linking this data with school records or other longitudinal data could provide insights on who the intervention is targeting most effectively.

Second, the LLX was designed as a self-directed learning intervention which specifically aimed to build children’s self-esteem, curiosity, soft skills (such as communication, critical

thinking, problem solving, public speaking), climate and gender sensitivity, and civic engagement. The assessment papers for LLX were confined to measuring academic performance on tested subjects. Research which can assess the broader development goals of the intervention would contribute considerably to our understanding of such interventions.

Looking ahead to the challenge of building back skills after a period of school closures, research on effective ways of engaging students with learning delays or age-grade mis-matches will be critical. Many aspects of the intervention studied here can also be applied to enrich the existing school curriculum, or can be delivered in intense summer camps to boost skills ahead of the school year. The experience of NGOs such as ASPIRE in this area could be further leveraged to test some of these elements in different settings.

Notes

1. Given the operational constraints present at the time, it was not possible for the NGO to follow a control group in a non-program area over the same period.
2. This comprehensive effort involves universalizing secondary education, using current research-based pedagogy, engaging local communities and administration, fostering teacher development, and leveraging technology for improved planning, monitoring, and enriched learning experiences. More information about this program can be found at: <https://aspire-india.org/projects/education-signature-program/>.
3. The NGO swiftly initiated the intervention in its program areas starting from April 2020, leveraging its established presence within these villages. At the onset of the pandemic, the NGO was officially recognized as a Covid-relief partner by the local government. Throughout the different pandemic-induced lockdowns, movement within villages was usually limited to government administration and local collaborators. As 98% of its staff members were locals, this official designation enabled the intervention to persist during most of this period.
4. Children from grades 1 and 2 were not included as they were considered too young to do the tasks on their own. Additionally, some very remote hamlets were unreachable due to logistical challenges. More details on selection into the LLX can be found in the the Supplementary Material.
5. The 2,700 volunteer teachers were mostly young people who were paid a daily honorarium of 125 INR or \$1.5 to cover the cost of their travel and data pack. For comparison, the minimum wage for a skilled worker in Odisha and Jharkhand is roughly \$5 per day.
6. As described in Section 2, schools opened for higher grades in October 2021 with plans to open for all grades soon. However, the Omicron wave hit soon after, leading to the closure of schools again. Due to budgetary constraints the LLX intervention did not continue beyond October 2021.
7. For more information on ASER, see <http://www.asercentre.org>. An English translation of the assessment papers is included in the Supplementary Materials (Section).
8. It is still likely that student recall, or teacher focus, are more likely to raise scores on questions with specific and memorable content. For example, upper primary students were asked about the capitals of various states in India, and also some questions about Covid-19. As a robustness check, we replicate our main analysis using only scores on math and language questions (see SM Table 4): the estimated effect sizes are smaller, but qualitatively similar.
9. Following their relative importance in the assessment paper, Writing and English were given double weight in the final score compared with the other sections.
10. This design counts each individual twice under different conditions, potentially conflating cross-sectional associations with heterogeneous treatment effects. In the Supplementary Material, we estimate the association between student characteristics and assessment scores split by pre and post time periods. While the trends are similar across time, the differences in assessment score by gender and social category are smaller after the intervention (SM Table 2).
11. When the interaction variable is binary, the fully interacted specification (which interacts our characteristic of interest with the post indicator, as well as all control variables including the constant) is equivalent statistically to running two separate regressions on samples split along the characteristic of interest. The interacted model, however, can be more straightforward to interpret. It also allows for interactions with continuous variables. These models include the variables in levels as well as interaction terms.
12. We further omit 42 observations from isolated panchayats: unlike other areas where the full block received the program, these are spread across 3 different blocks.
13. A recent randomized controlled trial evaluating a multi-faceted education intervention in the Gambia showed 3.2 SD improvements over 3 years (Eble et al. (2021)); while over a single year a local-language primary school program in Cameroon measured 1.44 SD improvements in overall test scores for first grade students (Laitin et al. (2019)).
14. Numbers shared by ASPIRE's finance department.
15. Equivalent figures for 2021 rupees are: 400-540 rupees per SD; 18.51-25.00 SD per 10,000 rupees.

Acknowledgment

The authors would like to thank the program staff at the NGO A Society for the Promotion of Inclusive and Relevant Education (ASPIRE), especially the learning vertical, who led the intervention during lockdown, for their advice, inputs, and tremendous effort in collecting the data. We thank Himangshu Kumar and Varun Bhattacharyya for research assistance. Special thanks to Dayaram, Executive Director at ASPIRE and Smita Agarwal, Education Head at Tata Steel Foundation, for their excellent contributions in enhancing our understanding of the intervention. Ethical approval has been granted by the University of St Andrews School of Economics & Finance Ethics Committee (approval code: EC16506). All views and errors are solely ours and this paper does not represent the view of ASPIRE or University of St Andrews, Scotland. Data availability statement: while the data used in this paper are confidential and property of ASPIRE, the NGO is open to discussing data access with bone-fide researchers. Interested researchers should email Kartika Bhatia to be put in touch with the relevant contacts at ASPIRE.

Disclosure statement

Kartika Bhatia was employed as Research Director at ASPIRE during part of this study.

References

- Andrew, A., & Salisbury, A. (2023). The educational experiences of Indian children during COVID-19. *Economics of Education Review*, 97, 102478. doi:10.1016/j.econedurev.2023.102478
- Angrist, N., Barros, A. D., Bhula, R., Chakera, S., Cummiskey, C., DeStefano, J., Floretta, J., Kaffenberger, M., Piper, B., & Stern, J. (2021). Building back better to avert a learning catastrophe: Estimating learning loss from COVID-19 school shutdowns in Africa and facilitating short-term and long-term learning recovery. *International Journal of Educational Development*, 84, 102397. doi:10.1016/j.ijedudev.2021.102397
- Angrist, N., Bergman, P., & Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour*, 6(7), 941–950. doi:10.1038/s41562-022-01381-z
- Bakhla, N., Drèze, J., Paikra, V., & Khera, R. (2021). *Locked out: Emergency report on school education*. India: School Survey.
- Brandt, W. C. (2020). *Measuring student success skills: A review of the literature on self-directed learning*. Dover, USA: Center for Assessment.
- Brussoni, M., Ishikawa, S., Brunelle, T., and., & Herrington, S. (2017). Landscapes for play: Effects of an intervention to promote nature-based risky play in early childhood centres. *Journal of Environmental Psychology*, 54, 139–150. doi:10.1016/j.jenvp.2017.11.001
- Crawford, L., Evans, D. K., Hares, S., & Sandefur, J. (2022). Live tutoring calls did not improve learning during the COVID-19 pandemic in Sierra Leone. *Journal of Development Economics*, 164, 103114. doi:10.1016/j.jdeveco.2023.103114
- Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Hsieh, P.-T. J., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., & Elbourne, D. (2021). How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia. *Journal of Development Economics*, 148, 102539. doi:10.1016/j.jdeveco.2020.102539
- Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17), 1–7. doi:10.1073/pnas.2022376118
- Government of India. (2020). *National Education Policy 2020*. New Delhi, India: Ministry of Human Resource Development.
- Grewenig, E., Lergetporer, P., Werner, K., Woessmann, L., & Zierow, L. (2021). COVID-19 and educational inequality: How school closures affect low- and high-achieving students. *European Economic Review*, 140, 103920. doi:10.1016/j.euroecorev.2021.103920
- Hassan, H., Islam, A., Siddique, A., & Wang, L. C. (2021, April). Telementoring and homeschooling during school closures: A randomized experiment in rural Bangladesh. *Munich Papers in Political Economy*, 1–21.
- Khan, M. J., & Ahmed, J. (2021). Child education in the time of pandemic: Learning loss and dropout. *Children and Youth Services Review*, 127, 106065. doi:10.1016/j.childyouth.2021.106065
- Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science (New York, NY)*, 340(6130), 297–300. doi:10.1126/science.1235350

- Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher*, 49(8), 549–565. doi:10.3102/0013189X20965918
- Laitin, D. D., Ramachandran, R., & Walter, S. L. (2019). The legacy of colonial language policies and their impact on student learning: Evidence from an experimental program in Cameroon. *Economic Development and Cultural Change*, 68(1), 239–272. doi:10.1086/700617
- Lee, R. L. T., Lane, S. J., Tang, A. C. Y., Leung, C., Kwok, S. W. H., Louie, L. H. T., Browne, G., & Chan, S. W. C. (2020). Effects of an unstructured free play and mindfulness intervention on wellbeing in kindergarten students. *International Journal of Environmental Research and Public Health*, 17, 139–150.
- Maldonado, J. E., & Witte, K. D. (2022). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*, 48(1), 49–94. doi:10.1002/berj.3754
- Radhakrishnan, K., Sabarwal, S., Sharma, U., Cullen, C., Crossley, C., Letsomo, T., & Angrist, N. (2021, July). Remote learning: Evidence from Nepal during COVID-19. *World Bank Policy Brief*, 1–8.
- Schueler, B. E., & Rodriguez-Segura, D. (2021). A cautionary tale of tutoring hard-to-reach students in Kenya. *Journal of Research on Educational Effectiveness*, 16(3), 442–472. doi:10.1080/19345747.2022.2131661
- The Economist. (2021, September 18). India's pupils have been hard hit by extended school closures. *The Economist*. https://www.economist.com/asia/2021/09/18/indias-pupils-have-been-hard-hit-by-extended-school-closures?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18156330227&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gclid=Cj0KCQjwsrWZBhC4ARIsAGGUJuqXrB-4LDYiH3k9xVn_-MKegibQn-MkyWfw5rOagFVmqV1VfENqgzUaAu93EALw_wcB&gclsrc=aw.ds
- UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. Paris, France: UNESCO.