Building Resilient Education Systems: Evidence from Large-Scale Randomized Trials in Five Countries

Noam Angrist, Micheal Ainomugisha, Saipramod Bathena, Peter Bergman, Colin Crossley, Claire Cullen, Thato Letsomo, Moitshepi Matsheng, Rene Marlon Panti, Shwetlena Sabarwal, Tim Sullivan*

April 14, 2023

Abstract

Education systems need to be able to withstand frequent shocks, including conflict, disease, natural disasters, floods, and climate events, all of which routinely shut down schools. During these shocks, alternative models are needed to deliver education. However, rigorous evaluation in these settings is challenging and rare, especially across multiple countries. We present results from large-scale randomized trials evaluating the provision of education in emergency settings across five countries: India, Kenya, Nepal, Philippines, and Uganda. We test multiple scalable models of remote instruction for primary school children during covid-19, which disrupted education for over 1 billion schoolchildren worldwide. Despite heterogeneous contexts, results show that the effectiveness of phone call tutorials can scale across contexts; we find consistently large and robust effect sizes on learning, with average effects of 0.30-0.35 standard deviations. These effects are highly cost-effective, delivering up to four years of high-quality instruction per $100 spent, ranking in the top percentile of all education programs and policies. In a subset of trials, we randomized whether the intervention was provided by NGO instructors or government teachers. Results show similar effects, indicating scalability within government systems. These results identify approaches to strengthen the resilience of education systems to withstand shocks, and more generally to deliver cost-effective education gains at scale.

.

Keywords: Human Capital, Education Systems, Education in Emergencies, Scale, Covid-19

*We are grateful to a coalition of partners who enabled this multi-country response. Implementing and research partners include: Youth Impact, J-PAL, Learning Collider, Oxford University, World Bank, Ministry of Education, Science and Technology of Nepal, Teach for Nepal, Street Child, Department of Education in the Philippines, IPA, Building Tomorrow, NewGlobe, Alokit, and Global School Leaders. Funding partners include UBS Optimus Foundation, Mulago Foundation, Douglas B. Marshall Foundation, Echidna Giving, SNF Foundation, Jacobs Foundation, JPAL Innovation in Government Initiative, and the World Bank.

I Introduction

More than 2 billion people live in countries affected by emergencies that frequently disrupt education. The causes of these disruptions are numerous and far-ranging, including rainy seasons, floods, pollution, elections, teacher strikes, conflict, climate shocks, disease, and natural disasters. For example, the spread of Ebola in West Africa in 2014 disrupted both education and health systems (Christensen et al. 2021), disrupting school for 1.7 million children for 9 months in Sierra Leone. Monsoon rains and flooding in Bangladesh, India, and Nepal from 2017 to 2019 closed 15,000 schools. Earthquakes in Pakistan, Haiti, and Nepal destroyed tens of thousands of schools.

While frequent and disruptive, education emergencies have historically been understudied. We present a new database documenting just how frequent and disruptive such shocks can be.[1] Schools close for lengthy periods during these emergencies, and learning loss can be substantial (Andrabi, Daniels, and Das 2021; Lichand et al. 2022; Carlana, La Ferrara, and Lopez 2023). Resilient education systems need to be able to withstand these shocks and continue to provide education. International aid organizations refer to these scenarios as "education in emergencies." Education Cannot Wait, the United Nation's Global Fund for education in emergencies, estimates that 222 million children are in active need of education in emergencies programs and policies.

Ideally, interventions and policies that can promote learning in emergencies are low cost, yield high take up across different geographies, and can be targeted to children of varying levels and cultural backgrounds. Understanding which approaches can be effective across these diverse circumstances requires multi-context, multi-model studies. However, rigorous evaluation of approaches to deliver education in emergencies remains challenging and rare. To date, most evaluations have been qualitative, with no multi-country experimental studies and few conducted with governments.

In this paper, we evaluate a set of education in emergencies programs to promote learning during large-scale school disruptions caused by COVID-19, which affected over 1 billion children worldwide. We conduct five randomized controlled trials – including delivery by NGOs and government teachers to test scalability within government systems – across five countries: India, Kenya, Nepal, Uganda, and the Philippines. In all settings, schooling was disrupted, and several of the countries in our study experienced some of the longest school closures in the world.

We use mobile phones to provide various educational interventions to primary school children. Mobile phones provide a platform that can cheaply reach students at scale in low-resource contexts. While less than 15 percent of households have access to the internet in low-income countries, over 70 percent have access to mobile phones (Carvalho and Crawford 2020). Moreover, mobile phones enable teachers to reach students at home even when school is disrupted, providing a resilient and

[1] Figure **??** documents the extent of school disruption for a selected set of emergencies over the last two decades. Notable examples include large earthquakes in Pakistan in 2005 and in Haiti in 2010, elections in Nepal in 2017 and air pollution in 2021, Ebola in Liberia and Sierra Leone, floods in Bangladesh in 2007 and Monsoon rains in 2019, prolonged droughts in Kenya in 2017, and foot and mouth disease in 2004 in Cambodia, among others. Conflicts in countries ranging from Afghanistan, Colombia, Ethiopia, Syria, and Myanmar have also been documented with substantial negative effects on education. For example, between 2015 and 2019 alone, more than 11,000 attacks on schools were documented in at least 93 countries (GCPEA, Education under Attack 2020).

flexible modality to provide education during emergencies. One treatment included a set of SMS messages, such as numeracy content provided weekly as well as nudges to engage in educational activities. A second treatment provided additional weekly 20 minute phone call tutorials for eight weeks. The educational pedagogy was as essential as the mobile phone platform. Phone calls covered foundational numeracy and aimed to target instruction to student learning levels via low-cost, high-frequency assessments. This approach builds on effective targeted approaches both in person and using technology (Banerjee et al. 2007; Banerjee et al. 2017; Muralidharan, Singh, and Ganimian 2019; Duflo, Kiessel, and Lucas 2020).

A proof of concept in Botswana showed phone call tutorials were effective in promoting learning during initial COVID-19 school disruptions (Angrist, Bergman, and Matsheng 2022). However, questions remain on whether this approach can be scaled across contexts and when delivered by governments – a pervasive challenge for social programs (List 2022; Mobarak 2022). This paper addresses the critical question of which types of education in emergency approaches can improve learning across a broad array of settings by conducting large randomized trials across five contexts as well as comparing scalable models, such as NGO and government delivery.

Our results show consistently large and robust effect sizes of phone call tutorials on learning across contexts, with average effects across all five countries of 0.30-0.35 standard deviations. We find results are largest in countries which experienced the longest school closures: Uganda and the Philippines. These results translate into large learning gains in absolute terms. In Uganda, for example, less than 20 percent of grade 4 students can divide at baseline, but by endline, nearly 50 percent can. These gains fully recover learning losses in math and enable substantial progress beyond status quo learning rates. On average, across all countries, we find a 65 percent increase in the percent of students who learn division. The effects are largest for students whose caregivers have only a primary education (rather than secondary and beyond), suggesting that results are strongest when there are fewer alternative educational support systems at home. Additional results show positive effects of phone call tutorials on learning higher-order competencies, such as fractions. Since fractions were not directly taught during the intervention, this provides evidence that learning extended beyond familiarity with the content taught. It further reveals dynamic complementarities, with the benefits of learning basic numeracy accruing to learning additional skills (Cunha and Heckman 2007).

We randomized whether the intervention was provided by NGO instructors or government teachers in the Philippines and Nepal. The average effect of phone call tutorials on learning when delivered by NGOs is 0.26 standard deviations and 0.31 when delivered by government teachers. These results show similar and statistically indistinguishable effects, indicating that government systems can effectively deliver these types of education in emergency responses. We also find high engagement across sites ranging from 70 to 80 percent, revealing the scalability and robustness of the approach even in disrupted and low-resource environments.

We further embedded a study randomly allocating a subset of government teachers to deliver phone call tutorials in Nepal. This experimental variation enables detection of the impact of

delivering the program on teacher beliefs and practices – an effect that could spillover into the education system that persist through teachers beyond the intervention. Results shows teacher practices shift substantially; teachers are 9.3 percentage points more likely to target their feedback to students learning level. Teachers are also more likely to get parents involved in education. We further find large effects on teacher perceptions that they were able to help students learn, as well as their desire to teach, with a 15.8 percentage point gain in wanting to be a teacher if they could make the choice again. These results suggest that delivering effective programs can unlock a virtuous cycle within government education systems, in turn motivating teachers to want to teach and to improve their teaching practice.

In contrast to phone call tutorials, the effects of SMS messages alone are mixed. On average, we find a 0.08 standard deviation effect on learning. While these effects are positive and statistically significant when pooled across contexts, they are not consistently statistically significant by country. Average effects are largely driven by substantial impacts in Uganda, with a 0.20 standard deviation effects that is significant at the 99 percent level, as well as effects in the Philippines, with an effect of 0.09 that is significant at the 90 percent level; there is no effect in Kenya or Nepal. These results suggest that SMS messages can work in contexts with the largest need, such as Uganda and the Philippines, but not in all contexts. Live phone call instruction, on the other hand, appears to best strike the balance of being intensive enough to deliver sustained impact across diverse contexts while remaining cheap and scalable.

Beyond the core set of randomized trials that occurred during COVID-19 school disruptions, an additional education emergency took place during the course of the study: a devastating typhoon in the Philippines that destroyed 4,000 classrooms and disrupted learning for 2 million children (OCHA 2022). The interaction between the areas affected by the typhoon and randomization to phone call tutorials and SMS messages resulted in a quasi-experiment. The results provide evidence on the disruptive effects of the typhoon on learning and approaches to stem learning losses in this additional education emergency context. Results show that the typhoon reduced learning by approximately 0.11-0.23 standard deviations. Despite the detrimental effects of the typhoon, phone call tutorials continued to be effective, with 0.26 standard deviation gains relative to the control group, although SMS messages alone were not enough to stem learning losses. These results reveal that phone call tutorial effectiveness persists across multiple education emergency settings.

We also contribute to the development of remote learning assessments. High-frequency remote assessment data enabled real-time targeting of instruction to student levels as well as the evaluation of program effectiveness. Even prior to the COVID-19 pandemic, phone calls have been used increasingly for household surveys, such as the World Bank Living Measurement Study (LSMS) or UNICEF's Multiple Indicator Cluster Survey (MICS). However, phone-based surveys are less frequently used to measure learning outcomes. We conduct five validity checks on high-frequency, low cost and scalable learning assessments via phone. A first check compares in-person to phone-based assessment for the exact same set of students in Kenya. We find no statistically distinguishable

difference between these two modes of assessments. An additional test included back-checks, with a random subset of students tested twice on the same competencies. We find a strong relationship as expected. We further randomize various problems of the same proficiency (e.g., four different questions to measure 2-digit addition with carryover). Results show no difference by question, showing accurate estimates of latent ability. Finally, we include a real-effort question to disentangle effects of the intervention on effort on the test, which has been shown to affect test scores during in-person exams (Gneezy et al. 2019), versus cognitive skills. We find no statistically significant effects of the interventions on effort, revealing that learning gains are indeed a function of cognitive skills.

In addition to assessing impact on learning outcomes, we examine impacts on beliefs and demand for the intervention. We find parents and children both update their beliefs broadly in line with true learning progress. These results build on a growing literature exploring the importance of parents knowing their child's learning level, enabling them to better support their education (Bergman 2021). In addition, caregivers demand the program. 96 percent of parents state they would like the phone call tutorials, which increases even further in phone call tutorial treatment groups, as does willingness to pay for the program. We also examine impacts on non-cognitive skills, such as perseverance and ambition, in line with Carlana and La Ferrara (2021). While we don't find statistically significant effects for SMS messages on their own, we find sizable effects on both of these outcomes in the phone call tutorials, with up to a 29 percent increase in ambition. We further find positive effects on measures of well-being, such as enjoying school and worrying less. A growing literature highlights the importance of both cognitive as well as non-cognitive skills for future life outcomes (Jackson 2018). These results provide experimental evidence that some educational interventions, such as the phone call tutorials, can promote both.

This study contributes to a nascent experimental literature on education in emergencies. Substantial research has taken place on this topic, however much of it has been qualitative or with small samples. One exception is a randomized trial where an NGO provided schooling in rural areas of conflict-affected regions in Afghanistan and found large effects on learning and closing of gender gaps (Burde and Linden 2013). We build on this literature by providing evidence from randomized controlled trials across multiple contexts and for government delivery models. We also expand the literature by evaluating alternative, scalable forms of education beyond traditional in-person schools, such as remote learning, which is the only option during many emergencies. In addition, a large literature documents the cost of school disruptions, such as teacher strikes (Jaume and Willen 2019), earthquakes (Andrabi et al. 2021), schools holidays (Cooper et al. 1996), and COVID-19 (Bacher-Hicks et al. 2021; Jack et al. 2021; Patrinos et al. 2022; Moscoviz and Evans 2022). However, less evidence exists on effective approaches to stem these learning losses. We contribute evidence on learning loss across multiple settings and, crucially, on scalable solutions to stem learning losses. This relates to an emerging evidence based on education interventions tested during COVID-19 (Carlana and La Ferrara 2021; Hassan et al. 2021; Crawfurd et al. 2021; Schueler and Rodriguez-Segura 2021; Lichand et al 2022; Hevia et al. 2022; Angrist et al. 2022).

We also contribute to a growing literature on scale. Recent examples show the extent of the

scaling challenge, with many social programs that initially worked in a proof-of-concept phase no longer delivering impact when scaled or delivered by governments (Mobarak 2022; List 2022).[2] In contrast with much of the existing literature, our results identify an education approach that *can* scale across contexts and within government systems. One reason might be that we leverage a particularly scalable technology: mobile phones. Few technologies have as widespread access across so many diverse contexts (Aker and Mbiti 2010; Aker et al. 2012). A second reason might be the generalizability of the underlying mechanisms tested in our study, which relate to other best practices in education. For example, tutoring has been shown to be one of the most effective, although expensive, approaches to improving learning in high-income settings (Nickow, Oreopoulos, and Quan, 2020; Robinson and Loeb 2021). The phone call tutorials in our study provide a cheap and scalable version of tutoring applicable in low- and middle-income contexts. In addition, since the phone calls are one-on-one and have frequent learning assessments, they enable highly targeted instruction to every child's learning level, another educational approach shown to consistently improve learning outcomes (Banerjee et al. 2017; Muralidharan, Singh, and Ganimian 2019; Duflo, Kiessel, and Lucas 2022).

In addition to identifying a scalable approach, our study advances the scale literature by conducting randomized trials across five countries in a literature where less than 1 percent of RCTs in a set of top economics journal articles are multi-country studies.[3] Some argue that while RCTs address internal validity concerns, they might not address important external validity questions (Pritchett and Sandefur 2015). This study highlights the potential for randomized trials to be conducted and coordinated across contexts – addressing both internal and external validity challenges simultaneously. While rare, this approach is gaining traction, with another prominent example including a multi-country study on microcredit (Banerjee et al. 2015). Finally, by evaluating the effectiveness of government delivery with experimental variation relative to NGO delivery across multiple contexts, we assess questions of scalability within government systems.

Third, we contribute to the global education literature, with a focus on improving learning outcomes. Over the past few decades education enrollments have improved worldwide, yet learning outcomes have barely budged (Pritchett 2013; World Bank 2018; Angrist et al. 2021). Estimates from our baseline survey, for example, show that 36 percent of students in grades 3 to 5 could not do any basic operations and only 7 percent of student could divide, falling well behind grade-level curriculum expectations. Growing evidence reveals that popular input-only reforms, such as gen-

[2]One example includes a contract teacher program in Kenya where effects dissipated when delivered by the government (Bold et al. 2018). Another example includes the diminishing effects of early childhood programs as they were scaled up from an efficacy trial ("proof of concept") in Jamaica, to a pilot in Colombia, to an at-scale program in Peru (Araujo, Rubio-Codina, and Schady 2021).

[3]Out of a set of 400 papers in development from 2019 and 2021 in a set of top economics journals, 19 percent were RCTs; of those that were RCTs about 1 percent of RCTs were multi-country studies. The set of journals considered includes the Top 5 economic journals (American Economic Review, Quarterly Journal of Economics, Econometrica, Journal of Political Economy, and Review of Economic Studies) and other top-tier general interest journals (Review of Economics and Statistics, Economic Journal, Journal of the European Economic Association, and all four American Economic Journal AEJ journals), and a top field journal (the Journal of Development Economics). Other prominent multi-country RCT efforts include the evaluations of Graduation programs (Banerjee et al. 2015) and Teaching at the Right Level (Banerjee et al. 2017) and some early grade reading interventions (Lucas et al. 2014).

eral teacher training, provision of computers, or school grants, are not enough to improve learning. In contrast, approaches that improve the quality of teaching, such as teaching at the right level and structured pedagogy, can generate large improvements in learning (Kremer, Brannen, and Glennerster 2013; Ganimian and Murnane 2016; Glewwe and Muralidharan 2016). The learning gains from the phone-based tutorials tested in this study can deliver up to 4 years of high-quality years of schooling per $100 spent. These effects rank in the top percentile of cost-effective interventions, benchmarked relative to 150 education policies and programs (Angrist et al. 2020). This reveals the potential of the approach to cost-effectively improve learning across low-and middle-income contexts and to help address a persistent global learning crisis.

The rest of this paper is structured as follows. Section II provides context and presents new data documenting how frequently education in emergencies occur. Section III describes the data collected across studies and sites. Section IV describes the experimental design and empirical strategy. Section V includes the results and cost-effectiveness analysis, and Section VI concludes.