Reading for Life: Lasting Impacts of a Literacy Intervention in Uganda

Julie Buhl-Wiggers, Jason T. Kerwin,
Ricardo Montero de la Piedra, Jeffrey Smith, and Rebecca Thornton*

PRELIMINARY AND INCOMPLETE—PLEASE DO NOT CIRCULATE

May 25, 2023

Abstract

The literature provides several examples of programs that affect educational outcomes in developing country contexts in the short run, but evidence of long-run effects remains scarce. We study the Northern Uganda Literacy Project (NULP)—an early grade reading intervention for children in grades one to three with large short-term impacts (1.3 SD in Leblango and 0.7 SD in English). We follow students eight years after the program began (and five years after it ended) and find 55% of the effect remains in Leblango and 79% remains in English. These effects represent 4.4 extra years of Leblango learning and 1.5 extra years of English learning compared to the control group. We find no spillover effects on math or sexual behavior. The control group exhibits dismal grade progression and retention both before and during the COVID-19 pandemic. While the NULP had no impact on attendance or remaining in school, it modestly improved grade progression.

JEL Codes: I2, O1

Keywords: Economics of Education, Human Capital, Development Economics, Long-run, Treatment Effects, Fade-out

*Buhl-Wiggers: Department of Economics, Copenhagen Business School (jubu.eco@cbs.dk); Kerwin: Department of Applied Economics, University of Minnesota and J-PAL (jkerwin@umn.edu); Montero de la Piedra: Department of Applied Economics, University of Minnesota (monte290@umn.edu); Smith: Department of Economics, University of Wisconsin (econjeff@ssc.wisc.edu) Thornton: Department of Economics, Baylor University (rebecca_thornton@baylor.edu). The randomized evaluation of the Northern Uganda Literacy Project would not have been possible without the collaboration of Victoria Brown and Ichuli Institute, Katherine Pollman, Deborah Amuka and other Mango Tree Educational Enterprises staff. We are grateful for funding from DFID/ESRC Raising Learning Outcomes Grant ES/M004996/2, Wellspring, and the International Growth Centre.

1 Introduction

A decade ago Lant Pritchett's book "The Rebirth of Education: Schooling Ain't Learning" laid out the fact that too many children leave school without basic skills and coined the term "learning crisis" to describe this global issue (Pritchett 2013). Since then, a plethora of studies have investigated how to address this crisis and help children learn (see e.g. reviews of the literature by Glewwe and Muralidharan 2016; Kremer, Brannen, and Glennerster 2013; Ganimian and Murnane 2014; Evans and Popova 2016). While some studies have found particular interventions to be effective in the short run, evidence of long-lasting effects remains scarce. In developing countries, only around 10 percent of RCTs that measure learning collect follow-up outcomes more than one month after the conclusion of the treatment (McEwan 2015). Moreover, the literature on the "fade-out" of the effects of educational interventions in the developed world, ably reviewed in Bailey et al. (2020), provides cause for concern about whether programs that are effective in the short run will meaningfully change the lives of the affected children in adulthood.

In this paper we study the Northern Uganda Literacy Project (NULP)—a highly effective early-grade reading intervention for children in grades one to three—and provide evidence of lasting learning gains five years after the program ended. The short-term impact (after three years of exposure) was 1.3 SDs in Leblango (the local language) and 0.7 SDs in English. These short-run impacts are incredibly large: they the NULP at the 99th percentile of all interventions that have been assessed via randomized trials in the developing world (Evans and Yuan 2022).

These large impacts lead to sustained gains. We find that 55% of the original effect remains for Leblango and 79% for English. This means that five years after the program ended, children in the treatment-group schools are still massively ahead in Leblango (by 0.71 SDs, equivalent to 4.4 additional years of schooling) and English (by 0.55 SDs, equivalent to 1.5 additional years of schooling). Despite these encouraging effects on reading, we find no spillover effects on math nor on sexual behavior outcomes. In addition, we find no evidence that the differences in learning stems from differences in the probability of working outside the home.

Even though the children in the original study should have reached secondary school (e.g., grade eight) by 2021, only a few of them have. We find that the vast majority of children in control group schools are more than three grades behind their expected level. The onset of COVID-19 only somewhat exacerbated this pre-existing pattern. The NULP mitigated this negative trajectory, with treatment group children experiencing about 10% less delay in grade progression than the control group. We can reject the null hypothesis of a zero effect on grade progression at conventional levels.

Our findings suggest that creating a solid foundation in early primary school can lead to sustained improvements in skills at the end of primary school. For these reading skills to transform into better life outcomes, either there need to be direct returns to reading on life outcomes such as in the labor or marriage market, or on measures of mental or physical health, or they need to increase the probability of attending secondary school and/or lead to better job opportunities. However, due to the massive grade retention—less than 1% of the children have progressed to secondary school—it remains too early to evaluate the NULP's effects on this outcome. Given that the children are in their early years of adolescence, we do not yet observe differences across treatment arms in other life outcomes.

Our study contributes to the limited literature on longer-run effects on educational interventions; see e.g. Bouguen et al. (2019) for a review of long-run impacts of RCTs in developing countries. Most current knowledge stems from studies of conditional cash transfers. From this body of work, the long-run effects on schooling are consistently positive, and there is some evidence of impacts on learning and labor market outcomes (Millán et al. 2019). Other studies focus on interventions providing school uniforms or scholarships. In Kenya, Evans and Ngatia (2021) do not find long-run effects of providing school uniforms in primary schools—even though school absenteeism decreased in the short run. In Ghana and Colombia, scholarships for secondary education increased school attainment and labor market participation, and also decreased teen fertility (Bettinger et al. 2019; Duflo, Dupas, and Kremer 2021). Another strand of literature evaluates school health or early-childhood development programs (e.g. Baird et al. 2016; Gertler et al. 2014). These interventions show effects on school participation as well as improved life outcomes 20 years later (Baird et al. 2016; Hamory et al. 2021; Gertler et al. 2014). However, only interventions aimed at very early ages seem to affect learning (Ozier 2018). While most of the existing knowledge base derives from interventions that increase the *quantity* of schooling through school attendance, our study evaluates an intervention aimed at improving the *quality* of schooling and finds relatively long-lasting effects on learning.

This paper proceeds as follows: First we set the stage and describe the NULP program in Section 2. In Section 3 we describe the original evaluation of the NULP as well as the sampling frame and long-term follow-up data collection. In Section 4 we describe our identification strategy and in Section 5 we present the results. Section 6 concludes.

2 Setting

2.1 Primary Education in Uganda

Primary education in Uganda consists of seven years of schooling (P1 to P7) and the official school starting age is six years. Since 1997, primary school has officially been free of charge, however, as resources are scarce many schools still depend on contributions from parents, thus *de facto* school fees are common and students whose parents are not able to meet these contributions are often sent home. The reform of 1997 was successful in enrolling children into school, especially at earlier grades; the net primary school enrollment rate is above 90% for both boys and girls (World Bank 2020). Yet, the large influx of children and limited resources has created raising concerns about diminishing school quality. Dropout and delayed progression through school remains an issue for both boys and girls.

The Ugandan national policy is to conduct all instruction in grades one to three (P1 to P3) using the local mother-tongue language. Grade four (P4) is a transition year, and grades five through seven (P5-P7) are taught in English. In grades one to three, students are taught reading and writing for an hour each day and take a half-hour of English class. The primary school curriculum includes a variety of teaching and learning materials such as syllabi, teachers' guides, resource books, and learning materials for students provided for by the government. Often, however the delivery and use of learning resources are often inadequate.

Teachers in Ugandan primary schools receive their basic teacher training at a Primary Teacher College (PTC). Once hired, teachers receive additional training and continuous professional development through the Teacher Development and Management System (TDMS). Under this system, Coordinating Centre Tutors (CCTs) conduct in-service teacher trainings and provide support and supervision of teachers through classroom monitoring visits that involve providing feedback and guidance on teaching quality and best practices. The TDMS often works with a cascade or "train-the-trainer" approach to training, intending that trainers pass on skills and competences to CCTs, who then directly train teachers.

The limited resources, low levels of teacher training and support, overall rates of poverty, and low levels of adult literacy throughout the country, coincides with severe educational problems in Uganda. After three years of completed primary education 41% of children still cannot read a single word, and by Grade 7 this number is still 10% (Uwezo 2016). To graduate from primary school, students must take the Primary Leaving Exam (PLE). Only 53% of students passed the PLE in 2017.

2.2 Northern Uganda Literacy Program

The Northern Uganda Literacy Project (NULP) was developed by Mango Tree Educational Enterprises Uganda, a private, locally-owned educational tools company that developed an instructional methodology to improve early-grade literacy, with a focus on mother-tongue-first instruction. In 2009, Mango Tree began developing their literacy program for one language group, Leblango. The NULP was piloted and refined from 2009 to 2012 to determine what was pedagogically and logistically successful. The project was based in the Lango sub-Region, where the vast majority of the population speaks Leblango.

Although the national policy is to instruct early primary students in their local language, in practice, with 41 different languages from three different language families, poorly developed orthographies, lack of teaching and reading materials, and deficient teacher training, English is still heavily used in primary schools across Uganda. Rote memorization of how to read basic words (in English) aloud is a common technique in early primary classrooms. The NULP model involved a revised curriculum for grades 1-3 that explicitly instructed teachers how to teach in the local language and avoided using any written English text during grade one. Moreover, NULP introduced content more slowly than the standard curriculum; approximately half of the letter/sounds of the alphabet to be taught in grade one, with the remaining taught in grade two.

To implement their program, the NULP provided extensive training and support for teachers in participating classrooms using expert trainers, detailed facilitator's guides, and instructional videos. The program provided teachers with a five-day residential (off-site) training in Leblango orthography and literacy prior to the beginning of the school year. Teachers also underwent two additional intensive, residential trainings on literacy methods during the breaks between the three school terms. In addition to the residential trainings, there were also six in-service training workshops on Saturdays throughout the school year. The training sessions were complemented by numerous support supervision visits that provided teachers with feedback about their teaching. These were conducted three times each term by Mango Tree staff members. In addition to the training and support supervision from Mango Tree staff, teachers also received two support supervision visits from CCTs each term. The CCTs were trained in providing the same sort of feedback on teaching performance as the Mango Tree staff, so each teacher received five total support supervision visits per term.

In addition to teacher training, NULP provided each classroom with a set of materials tailored for their instructional model including primers and readers. Teachers were provided with scripted lesson guides for each literacy lesson. Grade one classrooms were provided with slates that allowed each student to practice writing individually using pieces of chalk. Each classroom was provided with a wall clock to help teachers keep track of the time during a lesson.

Finally, NULP helped to support school parent meetings once per term and also actively engaged with communities more broadly to increase communities' knowledge of, and appreciation for, their local language. This included a radio program that conducted literacy and local-language promotion. The radio program and local-language promotion took place across the region and thus we cannot evaluate their effects in this paper.

3 Evaluation and Data

3.1 NULP Evaluation

To assess the impact of the NULP on student learning, we conducted a multi-year, randomized evaluation of the program. Schools were sampled for the study in two phases: an initial pilot RCT in 2013, and a larger RCT in 2014 which carried on in 2015 and 2016. In 2013, 38 eligible schools were selected to be part of the RCT. To be eligible schools had to meet a set of criteria established by Mango Tree, the most important being that each school needed to have exactly two P1 classrooms and teachers. In 2014 the program was expanded to 90 additional schools for a total of 128 schools. The eligibility criteria for these new 90 schools were slightly different, and less stringent.[1]

Of the 128 schools, the evaluation assigned each to one of three study arms: 1) Full-cost NULP, 2) Reduced-cost NULP, and 3) Control. In the Full-cost arm, schools received the original NULP as designed by and delivered by Mango Tree and its staff. In the Reduced-cost arm, some of the materials (slates and chalk) were eliminated, training was conducted through a cascade model led by government employees (Ministry of Education staff) rather than Mango Tree staff, and teacher received fewer support visits, again from government employees.[2] Schools in the Control group did not receive the literacy program. To randomize treatment assignments, schools were grouped into stratification cells of three schools each. Each stratification cell had its three schools randomly assigned to the three different study arms via a public lottery.

This study focuses on the cohort of students who entered the 128 study schools as first-graders in 2014.[3] The NULP was introduced to first-grade classrooms and teachers in treatment schools in 2014, and continued in 2015 and 2016 for second- and third-grade classrooms

[1] Criteria in 2014 include: having desks and blackboards in grade 1-3 classrooms and having a student-to-teacher ratio of no more than 150 students during the 2013 school year in grades P1 to P3.

[2] Kerwin and Thornton (2021) discuss and quantify the differences between the full- and reduced-cost program versions, and present the results of the 2013 pilot RCT.

[3] The treatment schools from the original 38 schools received the program in P1 in 2013, and then repeated the program in the same grade in 2014. For the other 90 schools, the program was first introduced in 2014.

and teachers.[4] Classrooms were allowed to keep all of the Mango Tree educational materials (such as slates, primers, and readers) after they received the program, but teachers no longer received additional training or support visits. This means that children starting Grade 1 in 2014 were treated for three years. To assess learning gains, 100 P1 students were sampled at random from each of the 128 schools.[5] This cohort was tracked into P2 and P3 in 2015 and 2016, respectively. Control-group schools received flipcharts for unrelated subjects in 2014, but no intervention after that point.

3.2 Long Term Follow-up

3.2.1 Sample and Data Collection

To assess the longer-term effects of the NULP we collected data for a targeted subset of the original cohort of students who were in P1 in 2014. The sample was designed to minimize attrition while maintaining a valid sampling frame balanced across study arms and we restricted the sample in the following ways. First, we selected only students who were sampled at baseline in 2014, excluding those who were added at the end of that school year. Second, we included only students who were randomly sampled for a household survey conducted in 2015, which gives us additional information critical for tracking the students. Third, we excluded all students from stratification cells that were in the city of Lira. This last exclusion was imposed because families from the city tend to move more, and thus tracking would be more challenging and costly in that area. Since the treatment assignment was randomized within each stratification cell, excluding these cells does not internally bias our estimates. After applying these criteria the target follow-up sample consisted of 3,098 students in 104 schools, which is slightly under half of the original sampled children in 2014 and four-fifths of the schools.

Prior to data collection, we held two school/community meetings at the end of 2020 to update information on the current location of the sampled children and mobilize parents for data collection. These efforts to locate the children was successful and only 10% of the sampled children were categorized as "Not found" at this stage. Of the 2,788 students who

[4] In 2017, Mango Tree piloted a teacher mentor program with fourth-grade teachers in the reduced-cost and full-cost schools to provide support; no materials or pedagogical training or support were delivered. This intervention was much less intensive than the earlier years.

[5] The sampling procedure differed slightly between the original 38 schools and the 90 schools added in 2014. In the 38 schools that participated in 2013, an initial sample of 40 grade one pupils was drawn at the 2014 baseline, and then 60 students were added at the 2014 endline following the same sampling procedure as at baseline. In the 90 new schools, 80 students were selected at baseline with an additional 20 added at endline. The difference was due to the organizational difficulty of testing large numbers of students at baseline or endline at each school, since the study also collected data on the second-grade students from the original 38 schools that had been exposed to the program in 2013.

could be located, 64% were still in the expected school/community and 26% were in a nearby school/community.

Due to the restrictions of COVID-19 the data collection process was divided into three phases; Phase 1: February 23rd to March 19th 2021 (grades 6 and 7)[6], Phase 2: October 18th to November 19th 2021 (grades 5 to 1), and Phase 3: May 30th to June 15th 2022 (students that could not be found in the two previous phases). We collected data from all 7 grades because there is a great delay in grade progression in our sample. According to official government policy, our cohort of students should have been in P7 in 2021, delayed one year relative to their initial trajectory due to the pandemic-related school closures. In fact, we find many students enrolled in far lower grades, as we discuss further below. We collected data on a total of 2,314 children; the attrition rate out of the overall targeted sample of 3,098 was 25.3%.

3.2.2 Data

We collected two types of data: learning assessments in Leblango, English, and math, and a student survey containing information on grade progression, school dropout, progression to secondary school, working outside of the home, and sexual behavior.

Reading skills in Leblango and English were assessed through the Early Grade Reading Assessment (EGRA), which is an internationally standardized exam designed to assess early literacy skills such as recognizing letters, reading simple words and understanding sentences and paragraphs. We use an adaptation of the EGRA to Leblango, which covers two components of literacy skills: oral reading fluency (ORF), and reading comprehension (RC). Math skills are measured through the Early Grade Math Assessment (EGMA), which covers three components of basic math skills; Addition, Subtraction and Numerical problems. In addition to the learning assessments collected in 2021 we also use the assessments collected at baseline in 2014.

To measure overall performance across all components of the assessments, we construct a principal components score index for each of the three subjects in the following way. For each assessment (English and Leblango EGRA, and EGMA) we estimate the weights of each skill in the first principal component for the control-group students in the 2014 baseline evaluation. Then we use those weights to predict this principal component among treated students in the baseline evaluations, as well as all students in the follow-up. We then standardize each index by subtracting the control mean and dividing by the control-group standard deviation in each year (2014 and 2021).

[6] As part of the gradual re-opening of school grades 7 and 6 opened up earlier than grade 5 and down.

School attendance is measured as a zero-one indicator where one indicates that the student ever attended school in a given year. School dropout is also measured as a zero-one indicator where a one indicates that the student ever dropped of school in a given year. Note that students often drop out of school and return within a single year, so these two indicators are not mutually exclusive. Grade progression measures which grade the child is attending in a given year. Attending secondary school is measured as a zero-one indicator where a one indicates that the child attended secondary school in a given year. Work outside the home is measured as a zero-one indicator where one indicates if the child worked outside the home in a given year. All of these measures are retrospective and asked of the child in the 2021 survey, for all years from 2014 through 2021.

Sexual experience is measured through three variables; ever had sex (where one indicates yes), age at sexual debut, and sexual debut before age 13 (where one indicates yes).

4 Empirical Strategy

Following our analysis plan, we obtain experimental impact estimates for each of our outcomes y_{ij} via the following parametric linear model estimated by ordinary least squares:

$$y_{ij} = \beta_0 + \beta_1 \text{FC}_j + \beta_2 \text{RC}_j + \mathbf{Z'}_j \tau + \mathbf{X'}_i \gamma + \varepsilon_{ijt} \quad (1)$$

where i indexes students, which are nested within their original schools (as of P1) indexed by j. FC_j and RC_j are indicators for a school being randomly assigned to the Full-cost or Reduced-cost program, respectively. Z_j is a vector of indicators for the stratification cells used in the lottery that assigned schools to study arms. X_i is a vector of control variables; we control for an indicator for being male, age as of the baseline exam (inputted as categorical indicators for each age)[7], and baseline test score indices for three exams (the Leblango EGRA, a math assessment, and an oral English exam). The baseline exams were conducted at the beginning of the 2014 school year. For students with missing values of the baseline exam score, we replace the missing values with zero and include a separate indicator variable for the baseline exam score being zero.

We conduct inference on our estimates via randomization inference. Specifically, we randomly permute the study arm assignments of each school within the stratification cells used in the original lottery. We implement this in Stata via the *ritest* command, following the approach in Kerwin and Thornton (2021).

We consider three null hypotheses for each outcome we study:

[7] We bottom-code age at 5 and top-code it at 9.

$$\mathrm{H}_0^1 : \beta_1 = 0 \tag{2}$$

$$\mathrm{H}_0^2 : \beta_2 = 0 \tag{3}$$

$$\mathrm{H}_0^3 : \beta_1 = \beta_2 \tag{4}$$

In our analysis plan we established three sets of outcomes we would study: confirmatory academic outcomes (Leblango EGRA, English EGRA, and EGMA scores in SD), confirmatory downstream outcomes (attended school in 2021 and attended secondary school in 2021), and exploratory outcomes (ever had sex, first had sex at age 13 or below, worked outside of the home in 2021, and worked outside of the home in a non-agricultural sector in 2021). In Section 5, we present the results for these outcomes, but we deviate from the plan in two ways. First, we adjust the way in which we construct EGRA and EGMA scores by only including the test components which were measured in the 2014 baseline and 2021 endline assessments. We do this to facilitate the estimation of treatment effects in equivalent years of schooling, which helps with the interpretation of the magnitude of the effects. Second, we changed the set of exploratory outcomes during the analysis process, as we discovered important results that were not considered in the analysis plan. The results that exactly follow the analysis plan are in Appendix A.2. The magnitudes and statistical significance are not meaningfully different from those presented in Section 5.

In Appendix A.2 we also take account of multiple hypothesis testing using the Benjamini, Krieger, and Yekutieli (2006) method to compute sharpened q-values that control the false discovery rate (FDR). Following Derksen et al. (2023), we use the Anderson (2008) implementation of their approach, which computes the lowest value of the sharpened q-value for which we can reject the null, so that our q-values can be interpreted in the same way that conventional p-values are[8]. We adjust for multiple testing separately by domain.

4.1 Attrition Bias

Given the relatively long time-span that our analysis covers and the attrition rates shown in Section 3, we assess if the program has any predictive power for attrition. We do this by regressing the attrition indicator (equal to 1 for students that attrited, 0 otherwise) on the two treatment indicators using the same specification we use to estimate treatment effects. We test this for the full sample of students who took the EGRA and EGMA tests, as well as for the subsample that also completed the survey that contains the information for the

[8] We thank Olivier Sterck for sharing the code used to estimate these q-values.

Table 1
Differential Attrition Test

	EGRA/EGMA	Additional Outcomes
Reduced-Cost	-0.009	-0.022
S.E.	(0.024)	(0.026)
R.I. p-value	[0.765]	[0.505]
Full-Cost	0.007	0.003
S.E.	(0.023)	(0.024)
R.I. p-value	[0.801]	[0.922]
Difference Between Treatments	0.016	0.025
S.E.	(0.025)	(0.026)
R.I. p-value	[0.669]	[0.507]
Control Mean	0.255	0.281
Control SD	0.436	0.450
N	3098	3098

Notes: Covariates are measured as of the baseline exams in 2014. All regressions control for stratification cell indicators and baseline values of Leblango EGRA Score, Oral English assesment, EGMA, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in brackets. Coefficients represent standard deviations of the control group.

confirmatory downstream outcomes. Although attrition rates are somewhat high (between 25.5% and 28.1% in the control group), Table 1 shows that there is no statistically significant difference between the attrition rates of the control group and either of the treatment arms.

We also use a more flexible approach by allowing for interaction between the treatment indicator and all the baseline covariates and joint testing for the statistical significance of all the interactions of each treatment arm. Table 2 shows that none of the sets of variables (baseline covariates in levels, the interactions with the reduced-cost treatment, and the interactions with the full-cost treatment) pass a joint significance test.

These results suggest that even though we were not able to follow about a fourth of the sample of students, the attrited children do not appear to be different than the rest in terms of our set of baseline covariates. Furthermore, neither of the treatments affects the probability of attriting from the sample. This implies that the risk of attrition causing bias in our estimates is low.

Table 2

Analysis of Differential Attrition with Interactions

	Levels	CCT Interaction	MT Interaction
Reduced-cost	-0.084		
	(0.137)		
Full-cost	0.069		
	(0.114)		
English EGRA score	0.028**	-0.043**	-0.005
	(0.013)	(0.020)	(0.019)
English EGRA score missing	-0.199*	-0.009	0.084
	(0.116)	(0.186)	(0.199)
Leblango EGRA score	-0.016	0.003	0.021
	(0.022)	(0.030)	(0.023)
Leblango EGRA score missing	0.074	0.044	-0.137
	-0.102	(0.143)	(0.152)
Math score	0.000	0.006	-0.023
	(0.019)	(0.025)	(0.025)
Math score missing	0.015	-0.023	0.015
	(0.030)	(0.049)	(0.045)
Male	0.008	0.038	0.04
	(0.031)	(0.038)	(0.037)
Age 5 or lower	-0.074	0.032	-0.055
	(0.098)	(0.136)	(0.141)
Age 6	-0.115	0.072	-0.094
	(0.072)	(0.131)	(0.108)
Age 7	-0.067	0.054	-0.103
	(0.070)	(0.127)	(0.111)
Age 8	-0.083	0.000	-0.072
	(0.075)	(0.128)	(0.113)
Age 9	-0.063	0.176	-0.108
	(0.090)	(0.137)	(0.121)
Joint p-value Interactions	[0.896]	[0.275]	[0.896]

5 Results

Table 3 presents the treatment effects of the full-cost and reduced-cost versions of the NULP on student test scores eight years after the program started and five years after it ended. We present these estimates in two forms. Columns (1), (3), and (5) measure the effects in SDs of the EGRA (English and Leblango) and EGMA (math) assessments. As mentioned in Section 4, we only include the assessment components that were common present in the baseline and endline evaluations. Columns (2), (4), and (6) present the results using a rescaled version of the dependent variable for the purpose of providing an impact coefficient that is easier to interpret. The coefficients in these columns show the impact of the program as measured in equivalent years of schooling (EYS), which expresses the learning gains in terms of how many additional years of schooling the control group would have to complete in order to catch up with the treatment group..

Table 3

Treatment Effects on Student EGRA (English and Leblango) and EGMA (Math) Scores

	English		Leblango		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
	SD	EYS	SD	EYS	SD	EYS
Reduced-cost	0.249***	0.686***	0.377***	2.355***	-0.036	-0.110
S.E.	(0.101)	(0.277)	(0.148)	(0.924)	(0.045)	(0.136)
R.I. p-value	[0.001]	[0.001]	[0.000]	[0.000]	[0.527]	[0.527]
Full-cost	0.547***	1.507***	0.712***	4.450***	-0.000	-0.001
S.E.	(0.141)	(0.387)	(0.200)	(1.248)	(0.041)	(0.125)
R.I. p-value	[0.001]	[0.001]	[0.000]	[0.000]	[0.996]	[0.996]
Difference Between Treatments	0.298	0.821	0.335	2.094	0.036	0.109
S.E.	(0.153)	(0.421)	(0.213)	(1.331)	(0.039)	(0.119)
R.I. p-value	[0.117]	[0.117]	[0.237]	[0.237]	[0.470]	[0.470]
Control Annual Gain (in SD)	0.363		0.160		0.330	
Control Mean	-0.000	-0.000	0.000	0.000	0.000	0.000
Control SD	1.000	2.755	1.000	6.250	1.000	3.030
N	2315	2315	2315	2315	2314	2314

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets [].

We find a large long-run impact of the NULP on student learning in English and Leblango, but not in math. Five years after the intervention ended, students who received the reduced-cost version of the program scored 0.25 and 0.38 SDs higher than the control students on the English and Leblango EGRA, respectively. More impressively, the full-cost version of the program had sustained impacts on student learning of 0.55 and 0.71 SDs, respectively. The math estimates show a small and statistically insignificant negative impact of -0.04 SDs for the reduced-cost version and a point estimate of exactly zero (to three decimal places) for the full-cost version. We can reject impacts on math larger than 0.08 SDs, which is smaller than the average effect size of 0.1 SDs for education RCTs in the developing world (Evans and Yuan 2022).

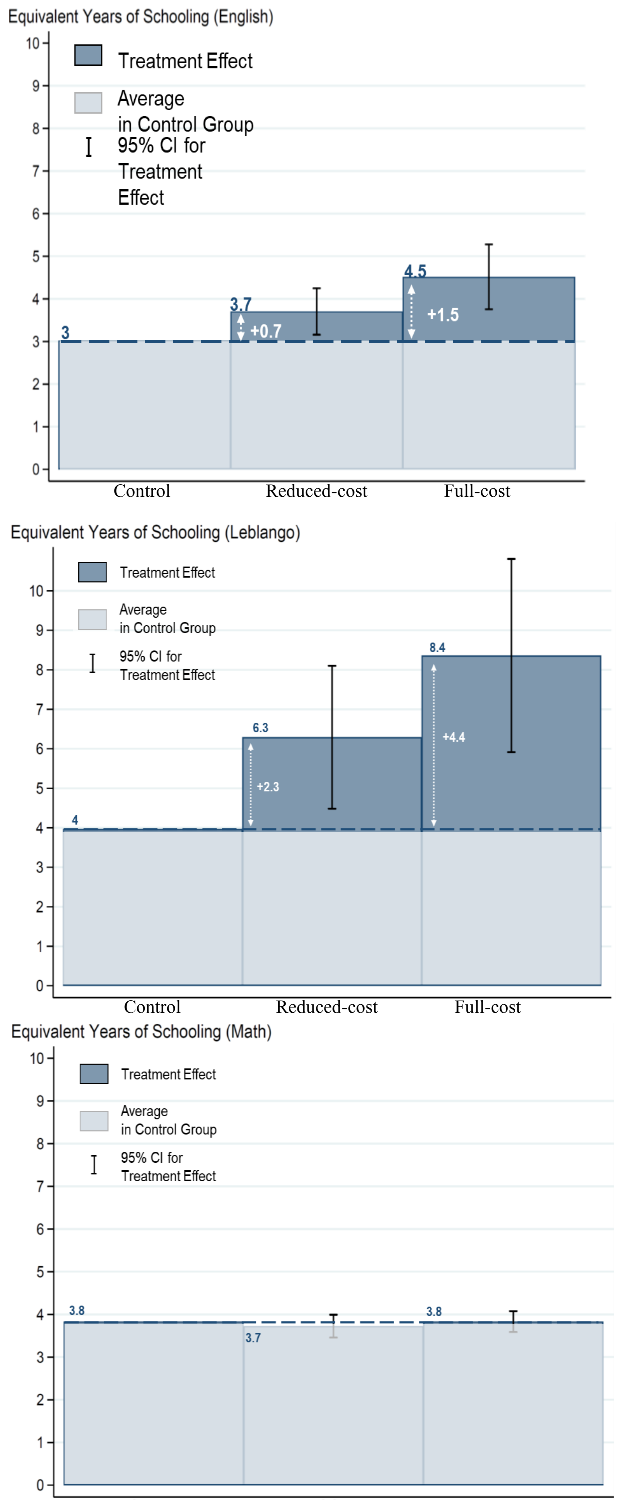
The magnitude of these gains may be easier to appreciate in terms of equivalent years of schooling. Figure 1 shows the average number of actual completed years of schooling for control-group students as well as the learning gains caused by the treatment measured in years of schooling. Completed years of schooling for the control group are measured since the earliest test score available for each subject (the beginning of 2014 for Leblango and math and the end of 2015 for English). Note that even though eight years had elapsed from the beginning of intervention (from 2014 to 2021), the number of *completed grades* is far less than eight. The average student in the control group had completed just three grades since the end of 2015 despite six school years having elapsed since then.

For the English EGRA, control-group students had progressed 3 grades between tests on average. The students in the reduced-cost arm accumulated learning equivalent to an additional 0.7 years, while the students in the full-cost arm gained 1.5 additional years, compared to the control group. For Leblango, the control-group students had progressed 4 grades on average between tests. Students who received the reduced-cost treatment gained an additional 2.3 years of learning, and students in the full-cost treatment arm gained 4.4 more years of learning. Finally, the average control student that took the EGMA (math) completed 3.8 years of schooling, while students on both versions of the program had similar learning progress to the control group. Although the estimates for the full-cost version of the program are higher than those for the reduced-cost version in all cases, we cannot reject the null hypothesis that the gains are equal.

Table 4 shows the effect of the NULP on self-reported school attendance in 2019, 2020, and 2021. Almost all control-group students report attending school during 2019 and 2020 (95% and 94%, respectively), but attendance dropped to 84% in 2021 due to the COVID-19 pandemic. Reported attendance in 2021 is surprisingly high considering that all government schools in Uganda were closed for the vast majority of that year. This leads us to believe that students reported attending even if they actually only went to their schools for a small number of days those years.

We find no statistically significant effect of the NULP on attending school in any of the three years. The point estimates for the reduced-cost version of the program were 0.7, 1.3 and 1.6 percentage points for 2019, 2020 and 2021 respectively, while the full-cost point estimates were -1.3, -0.5 and 0.1 percentage points for the same years. In these analyses we have 80% power to detect effects of 3.1 percentage points in 2019 and 2020 and 6.3 percentage points

Figure 1
Treatment Effects on EGRA and EGMA in EYS



Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Whiskers represent heteroskedasticity-robust 95% confidence intervals, clustered by school. Detailed results are presented in Table 3.

Table 4
Treatment Effects on School Attendance

	2019	2020	2021
Reduced-cost	0.007	0.013	0.016
S.E.	(0.010)	(0.011)	(0.025)
R.I. p-value	[0.599]	[0.367]	[0.575]
Full-cost	-0.013	-0.005	0.001
S.E.	(0.011)	(0.011)	(0.024)
R.I. p-value	[0.344]	[0.743]	[0.975]
Difference Between Treatments	-0.020	-0.017	-0.015
S.E.	(0.010)	(0.011)	(0.024)
R.I. p-value	[0.141]	[0.196]	[0.645]
Control Mean	0.945	0.937	0.836
Control SD	0.228	0.244	0.371
N	2252	2251	2249

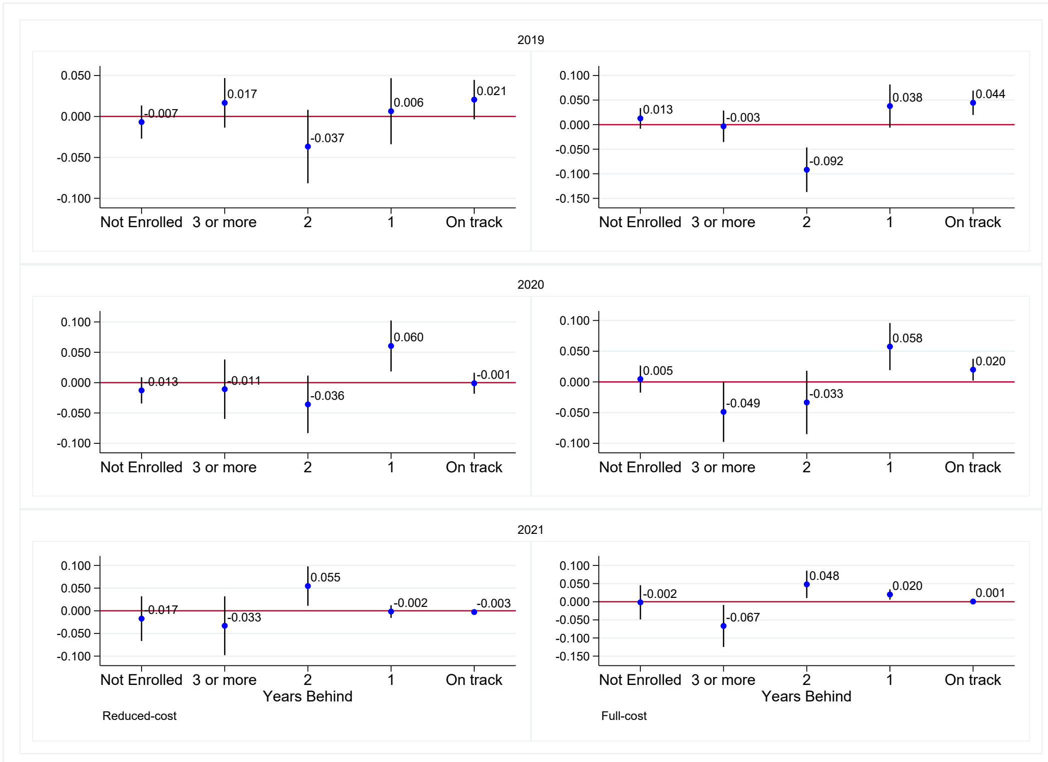
Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference *p*-values, clustered by school (104) and stratified by stratification cell (36), in square brackets [].

in 2021. We thus cannot rule out the possibility of small effects on attendance in 2021.

We originally planned to estimate the effects of the NULP on the probability of attending secondary school. However, we found only four students who managed to progress to that level (two in the control group and two in the full-cost version of the program). This was due to the Ugandan government's response to the COVID-19 pandemic, which delayed all grade progression by one year; the tiny number of students observed to be in grade 8 are likely due to measurement error. The fact that grade progression was delayed means that the effect of the NULP on advancing to grade 8 is mechanically zero. Therefore, we use an alternative approach to evaluate whether the program changed students' grade progression.

Figure 2 (and Appendix Table A1) shows the effect of the NULP on keeping on track with the grade that the student should be on if they progressed one grade level for every school year since 2014. This would mean that we would expect students to be in grade 8 by 2021 (which, as detailed above, essentially never happened). We find that the program is significantly reduces delay in grade progression. In 2019, the full-cost version of the program reduced the probability of being two years behind by 9.2 percentage points, and correspondingly increased the share of students who were one year behind or on track (although the increases are not statistically significant under randomization inference). We see a similar qualitative pattern for the reduced-cost program, but the effects are not statistically significant. In 2020, the

Figure 2
Treatment Effects on Grade Progression



Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Whiskers represent heteroskedasticity-robust 95% confidence intervals, clustered by school. Detailed results are presented in Table A1.

reduced-cost version of the program increased the probability of being one year behind by 6 percentage points, reducing the probability of being further behind. The full-cost version of the program increased the probability of being one year behind by 5.8 percentage points and the probability of being on track by 2.0 percentage points, by reducing the probability of being two or three years behind. Finally, by 2021, the full-cost version of the program reduces the probability of being three or more years behind by 6.7 percentage points and increasing the probability of being two years behind by 4.8 percentage points and of being one year behind by 2.0 percentage points.

We also explored additional potential outcomes that could be related to the impact that the program has on the educational outcomes. In particular, we focused on sexual behaviors and labor market participation. Table 5 shows the effects of the program on sexual behavior. By 2021, about 11% of the control students reported had had sex at least once. We find no evidence that either version of the program had any effect on the probability of ever having sex. Among the control students who had had sex, the average age when students had

Table 5
Treatment Effects on Sexual Behavior

	Ever Had Sex	Age First Had Sex	Had Sex at 13 or Before
Reduced-cost	-0.010	-0.237	-0.011
S.E.	(0.020)	(0.438)	(0.008)
R.I. p-value	[0.714]	[0.645]	[0.297]
Full-cost	0.006	-0.252	-0.008
S.E.	(0.019)	(0.501)	(0.010)
R.I. p-value	[0.788]	[0.656]	[0.541]
Difference Between Treatments	0.016	-0.015	0.002
S.E.	(0.018)	(0.504)	(0.008)
R.I. p-value	[0.471]	[0.969]	[0.802]
Control Mean	0.109	13.923	0.037
Control SD	0.312	1.929	0.189
N	1690	148	1690

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets [].

their first sexual experience was about 14 years old. Again, we did not find any evidence of the program having any effect on this variable. Finally, we examined whether the program changed the probability of having sex before the age of 13. Just under 4 percent of the control group had had sex prior to age 13, but we find no statistically significant effects of the NULP on this outcome.

We also explored the possibility that the program affected the probability of the children having a job outside of home. We find that in 2019, 39% of the control students worked outside of their homes. This percentage grew to 50% and 54% in 2020 and 2021, respectively. Neither version of the program had a statistically significant effect on the probability of working outside of home in any of the three years.

Overall, these results suggest that while the NULP had an impressive and long-lasting effect on Leblango and English learning, as well as a moderate effect in grade progression, these gains did not translate into math learning or other behaviors.

Because our analysis in this paper deviates from our original analysis plan, we present the results based on exactly what we pre-specified in Appendix Section A.2. These results also show the (Anderson 2008) FDR-adjusted q-values for the pre-specified outcomes. Appendix Table A2 differs from our main results Table 3 in the exact way that we calculate the test score indices. For our main analysis, we compute the score indices using only those test score components that were available at both baseline and endline, so that our scores can

Table 6
Treatment Effects on Working Outside of Home

	2019	2020	2021
Reduced-cost	-0.008	0.006	-0.003
S.E.	(0.020)	(0.022)	(0.022)
R.I. p-value	[0.710]	[0.846]	[0.926]
Full-cost	-0.044	-0.030	-0.027
S.E.	(0.023)	(0.024)	(0.020)
R.I. p-value	[0.125]	[0.346]	[0.237]
Difference Between Treatments	-0.036	-0.036	-0.025
S.E.	(0.022)	(0.022)	(0.021)
R.I. p-value	[0.220]	[0.221]	[0.373]
Control Mean	0.393	0.501	0.543
Control SD	0.489	0.500	0.498
N	2228	2231	2227

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assesment, EGMA, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets [].

be measured in equivalent years of schooling. In the analysis plan, we pre-specified that we would use all components available at endline. This change makes no difference for our substantive conclusions about the program's effects on test scores. The significance of our estimated treatment effects on test scores is also robust to adjusting for multiple testing. This pattern carries over to Appendix Table A3 and Appendix Table A4: if we run the analyses exactly as pre-specified in our analysis plan, we find no statistically significant effects of the NULP on attending school, attending secondary school, sexual behavior, and working outside the home.

6 Conclusion

Our findings from this study of the Northern Uganda Literacy Project (NULP) shed light on the sustained effects of education interventions in developing countries. While the literature on such programs primarily focuses on short-term outcomes, this research demonstrates that the NULP had a significant and lasting impact on early grade reading skills, with positive spillovers onto English. Eight years after the program began, a substantial portion of the effects persisted. The study also revealed that the NULP did not have spillover effects onto math or sexual behavior, but did cause improvements in grade progression.

Our findings also highlight dismal overall grade progression in Uganda, even before delays due to school shutdowns during COVID-19. There are limited longitudinal studies that document grade progression (or lack of) in developing countries,[9] and this is an area of important future research.

These findings contribute to the existing literature by highlighting the potential of well-designed early-grade educational interventions for enhancing educational outcomes in developing countries. The sustained effects observed in this study emphasize the importance of sustained efforts and investments in early education programs to support children's learning trajectories—particularly in literacy. Further research is warranted to explore the mechanisms and factors that contribute to the durability of the NULP's effects, as well as to investigate ways to extend its benefits to other academic areas and broader aspects of students' lives.

[9] One exception is Lam, Ardington, and Leibbrandt (2011), which documents very poor grade progression for black students in South Africa.

References

Anderson, Michael L (2008). "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". *Journal of the American statistical Association* 103.484, pp. 1481–1495.

Bailey, Drew H., Greg J. Duncan, Flávio Cunha, Barbara R. Foorman, and David S. Yeager (2020). "Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions". *Psychological Science in the Public Interest* 21.2. Publisher: SAGE Publications Inc, pp. 55–97. ISSN: 1529-1006. DOI: 10.1177/1529100620915848.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel (2016). "Worms at Work: Long-run Impacts of a Child Health Investment". *The Quarterly Journal of Economics* 131.4, pp. 1637–1680. ISSN: 0033-5533. DOI: 10.1093/qje/qjw022.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli (2006). "Adaptive linear step-up procedures that control the false discovery rate". *Biometrika* 93.3, pp. 491–507.

Bettinger, Eric, Michael Kremer, Maurice Kugler, Carlos Medina, Christian Posso, Juan E Saavedra, et al. (2019). *School vouchers, labor markets and vocational education*. Banco de la Republica Colombia.

Bouguen, Adrien, Yue Huang, Michael Kremer, and Edward Miguel (2019). "Using Randomized Controlled Trials to Estimate Long-Run Impacts in Development Economics". *Annual Review of Economics* 11.1. _eprint: https://doi.org/10.1146/annurev-economics-080218-030333, pp. 523–561. DOI: 10.1146/annurev-economics-080218-030333.

Derksen, Laura, Jason T Kerwin, Natalia Ordaz Reynoso, and Olivier Sterck (2023). "Healthcare Appointments as Commitment Devices".

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2021). *The Impact of Free Secondary Education: Experimental Evidence from Ghana*. Working Paper 28937. Series: Working Paper Series. National Bureau of Economic Research. DOI: 10.3386/w28937.

Evans, David K and Mũthoni Ngatia (2021). "School uniforms, short-run participation, and long-run outcomes: Evidence from Kenya". *The World Bank Economic Review* 35.3. Publisher: Oxford University Press, pp. 705–719.

Evans, David K. and Anna Popova (2016). "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews". *The World Bank Research Observer* 31.2, pp. 242–270. DOI: 10.1093/wbro/lkw004.

Evans, David K. and Fei Yuan (2022). "How Big Are Effect Sizes in International Education Studies?" en. *Educational Evaluation and Policy Analysis* 44.3. Publisher: American Educational Research Association, pp. 532–540. ISSN: 0162-3737. DOI: 10.3102/01623737221079646.

Ganimian, Alejandro J. and Richard J. Murnane (2014). *Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Impact Evaluations*. Working Paper 20284. National Bureau of Economic Research. DOI: 10.3386/w20284.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeerch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor (2014). "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica". *Science (New York, N.Y.)* 344.6187, pp. 998–1001. ISSN: 0036-8075. DOI: 10.1126/science.1251178.

Glewwe, Paul and Karthik Muralidharan (2016). "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications". *Handbook of the*

Economics of Education. Ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 5. Elsevier, pp. 653–743. DOI: 10.1016/B978-0-444-63459-7.00010-5.

Hamory, Joan, Edward Miguel, Michael Walker, Michael Kremer, and Sarah Baird (2021). "Twenty-year economic impacts of deworming". *Proceedings of the National Academy of Sciences* 118.14. Publisher: Proceedings of the National Academy of Sciences, e2023185118. DOI: 10.1073/pnas.2023185118.

Kerwin, Jason T. and Rebecca L. Thornton (2021). "Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures". *The Review of Economics and Statistics* 103.2, pp. 251–264. DOI: 10.1162/rest_a_00911.

Kremer, Michael, Conner Brannen, and Rachel Glennerster (2013). "The Challenge of Education and Learning in the Developing World". *Science* 340.6130, pp. 297–300. DOI: 10.1126/science.1235350.

Lam, David, Cally Ardington, and Murray Leibbrandt (2011). "Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa". en. *Journal of Development Economics* 95.2, pp. 121–136. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2010.05.005.

McEwan, Patrick J. (2015). "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments". *Review of Educational Research* 85.3, pp. 353–394. DOI: 10.3102/0034654314553127.

Millán, Teresa Molina, Tania Barham, Karen Macours, John A. Maluccio, and Marco Stampini (2019). "Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence". *World Bank Research Observer* 34.1. Publisher: Oxford University Press, pp. 119–159. ISSN: 1564-6971. DOI: 10.1093/wbro/lky005.

Ozier, Owen (2018). "Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming". *American Economic Journal: Applied Economics* 10.3, pp. 235–262. ISSN: 1945-7782. DOI: 10.1257/app.20160183.

Pritchett, Lant (2013). *The rebirth of education: Schooling ain't learning*. CGD Books.

Uwezo (2016). *Are Our Children Learning (2016)? Uwezo Uganda Sixth Learning Assessment Report*. Kampala: Twaweza East Africa.

World Bank (2020). "School enrollment, primary (% net)". *World Development Indicators*.

A Online Appendix

A.1 Appendix Tables

Appendix Table A1

Treatment Effects on Grade Progression

	Not Enrolled	3 or More Years Behind	2 Years Behind	1 Year Behind	On Track
2019					
Reduced-cost	-0.007	0.017	-0.037	0.006	0.021
S.E.	(0.010)	(0.015)	(0.023)	(0.020)	(0.012)
R.I. p-value	[0.592]	[0.390]	[0.213]	[0.798]	[0.182]
Full-cost	0.013	-0.003	-0.092***	0.038	0.044
S.E.	(0.011)	(0.016)	(0.023)	(0.022)	(0.013)
R.I. p-value	[0.343]	[0.883]	[0.001]	[0.185]	[0.005]
Difference	0.020	-0.020	-0.055**	0.031	0.024
S.E.	(0.010)	(0.017)	(0.024)	(0.024)	(0.014)
R.I. p-value	[0.140]	[0.360]	[0.047]	[0.292]	[0.239]
Control Mean	0.055	0.117	0.474	0.304	0.050
Control SD	0.228	0.322	0.500	0.460	0.217
N	2251	2251	2251	2251	2251
2020					
Reduced-cost	-0.013	-0.011	-0.036	0.060**	-0.001
S.E.	(0.011)	(0.025)	(0.024)	(0.021)	(0.009)
R.I. p-value	[0.367]	[0.746]	[0.249]	[0.036]	[0.921]
Full-cost	0.005	-0.049	-0.033	0.058**	0.020*
S.E.	(0.011)	(0.025)	(0.026)	(0.019)	(0.009)
R.I. p-value	[0.743]	[0.146]	[0.316]	[0.013]	[0.080]
Difference	0.017	-0.038	0.002	-0.003	0.021*
S.E.	(0.011)	(0.026)	(0.025)	(0.020)	(0.008)
R.I. p-value	[0.196]	[0.286]	[0.941]	[0.902]	[0.063]
Control Mean	0.063	0.358	0.405	0.149	0.025
Control SD	0.244	0.480	0.491	0.356	0.156
N	2251	2251	2251	2251	2251
2021					
Reduced-cost	-0.017	-0.033	0.055	-0.002	-0.003
S.E.	(0.025)	(0.033)	(0.022)	(0.007)	(0.002)
R.I. p-value	[0.555]	[0.436]	[0.062]	[0.862]	[0.425]
Full-cost	-0.002	-0.067*	0.048**	0.020**	0.001
S.E.	(0.024)	(0.029)	(0.019)	(0.007)	(0.002)
R.I. p-value	[0.963]	[0.075]	[0.043]	[0.037]	[0.926]
Difference	0.016	-0.034	-0.007	0.022**	0.003
S.E.	(0.024)	(0.029)	(0.020)	(0.007)	(0.002)
R.I. p-value	[0.644]	[0.374]	[0.830]	[0.024]	[0.187]
Control Mean	0.165	0.646	0.160	0.026	0.003
Control SD	0.372	0.479	0.367	0.160	0.053
N	2235	2235	2235	2235	2235

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets [].

A.2 Analysis Plan Results

This appendix shows the estimations of the treatment effects of the NULP that mirror the preanalysis plan. These results don't change the conclusions presented in Section 5.

Appendix Table A2

Treatment Effects on Confirmatory Academic Outcomes

	English	Leblango	Math
Reduced-cost	0.244***	0.377***	-0.027
S.E.	(0.099)	(0.148)	(0.048)
R.I. p-value	[0.001]	[0.000]	[0.662]
q-value	{0.002}	{0.001}	{0.284}
Full-cost	0.537***	0.712***	0.002
S.E.	(0.138)	(0.200)	(0.044)
R.I. p-value	[0.001]	[0.000]	[0.970]
q-value	{0.002}	{0.001}	{0.478}
Difference Between Treatments	0.293	0.335	0.028
S.E.	(0.150)	(0.213)	(0.043)
R.I. p-value	[0.118]	[0.237]	[0.589]
q-value	{0.548}	{0.548}	{0.552}
Control Mean	-0.000	-0.000	-0.000
Control SD	1.000	1.000	1.000
N	3098	3098	3098

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets []. Multiple testing adjusted q-values in curly braces {}. Coefficients represent standard deviations of the control group.

Appendix Table A3

Treatment Effects on Confirmatory Downstream Outcomes

	Attended School (2021)	Attended Secondary School (2021)
Reduced-cost	0.016	-0.003
S.E.	(0.025)	(0.002)
R.I. p-value	[0.575]	[0.424]
q-value	{0.935}	{0.935}
Full-cost	0.001	0.001
S.E.	(0.024)	(0.002)
R.I. p-value	[0.975]	[0.925]
q-value	{0.935}	{0.935}
Difference Between Treatments	-0.015	0.003
S.E.	(0.024)	(0.002)
R.I. p-value	[0.645]	[0.176]
q-value	{0.583}	{0.583}
Control Mean	0.836	0.003
Control SD	0.371	0.052
N	2249	2249

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets []. Multiple testing adjusted q-values in curly braces {}.

Appendix Table A4

Treatment Effects on Exploratory Outcomes

	Ever had Sex	First had sex at age 13 or below	Worked outside of the home
CCT	-0.010	-0.011	-0.003
S.E.	(0.020)	(0.008)	(0.022)
R.I. p-value	[0.714]	[0.297]	[0.926]
q-value	{1.000}	{1.000}	{1.000}
MT	0.006	-0.008	-0.027
S.E.	(0.019)	(0.010)	(0.020)
R.I. p-value	[0.788]	[0.541]	[0.237]
q-value	{1.000}	{1.000}	{1.000}
Difference Between Treatments	0.016	0.002	-0.025
S.E.	(0.018)	(0.008)	(0.021)
R.I. p-value	[0.471]	[0.802]	[0.373]
q-value	{1.000}	{1.000}	{1.000}
Control Mean	0.109	0.037	0.543
Control SD	0.312	0.189	0.498
N	1690	1690	2227

Notes: All regressions control for stratification cell indicators and baseline values of the Leblango EGRA Score, Oral English assessment, math test, gender and age; missing values of control variables are dummied out. Heteroskedasticity-robust standard errors, clustered by school, in parentheses (). Randomization inference p-values, clustered by school (104) and stratified by stratification cell (36), in square brackets []. Multiple testing adjusted q-values in curly braces {}.