Schools, Staffing, and Free Lunch (No, not that kind)

Wyatt Brooks[1], Viva Ona Bartkus[2], Joseph Kaboski[2],
Carolyn Pelnik[3], and Maurice Sikenyi[2]

[1]Arizona State University
[2]University of Notre Dame
[3]World Bank Development Research Group

February 2026

Abstract

Education in developing countries is often plagued by the twin challenges of low school quality and tight educational budgets. Teachers and principals are constrained in how they spend their time, and staffing may not be optimal. To explore cost-effective personnel changes, we conduct a field experiment in 112 rural Ugandan schools where we shock the labor structure of the schools by exogenously adding staff of different types. We assess students on numeracy and literacy and measure impacts on the time allocation across tasks completed by different types of workers (teachers and administrators) within schools. We find that adding a new layer of management together with new workers significantly improves student numeracy by one-third of a standard deviation and results in the reallocation of time between tasks by existing workers. Using the experimental data, we estimate a task-based production function for schools in which workers of various types differ in their comparative advantage across tasks and optimally allocate their time between tasks in order to maximize student learning. Using the model, we can predict how student outcomes change with counterfactual worker composition. We find a free lunch: counterfactual cost-neutral personnel changes can lead to large improvements in student outcomes, roughly one-quarter of a standard deviation in test scores for the median school.

1 Introduction

The lagging quality of schools in poor countries is a well-known bottleneck to development of both macroeconomies and individuals.[1] Educational budgets are strained, however, so financially-strapped school systems need to look for ways to utilize existing resources more effectively. Competition and the profit motive, the main forces we assume will push us toward organizational efficiency in the private sector, are absent in the public sector. Are there gains to be had from re-organizing existing resources?

This paper uses a randomized controlled trial (RCT) of human resource injections together with a structural model to identify a managerial free lunch: a reallocation of resources that is cost-neutral and improves school quality. We focus on the roles of relatively expensive, highly educated managerial staff versus less expensive (but also less trained) workers and find that the two interact in important ways to improve school quality. Specifically, we randomly allocate new staff of these two types to government-run schools in rural Uganda and find that adding new workers together with a layer of highly educated management increases student numeracy by twice as much adding workers alone. The presence of new staff also leads existing workers to reallocate their time between tasks. We use the results of this experiment to estimate a task-based production function for schools in which workers of various types differ in their comparative advantage across tasks and optimally allocate their time between tasks to maximize student learning. Our model shows how large improvements in student outcomes can result from cost-neutral changes in labor structure.

We partner with Building Tomorrow (BT), a Ugandan-American non-governmental organization (NGO), to evaluate two different programs that add staff of various types to rural, government-run schools. In the first program, the NGO implements a highly competitive search process to select and train college-educated "Fellows" who are deployed to schools to help classroom teachers and head teachers (akin to school principals) address the challenges facing schools. Fellows also recruit community volunteers to implement remedial education. In this model, schools receive a new layer of highly educated management (the Fellows) alongside new workers (the Community Education Volunteers, or CEVs). In the second program, only new workers (the CEVs) offer supplementary remedial education, without a manager. In this setting, the cost of adding one manager is more than sixty times the cost of adding one worker, consistent with the extremely high relative cost of managers in low-income countries, as documented by Hjort et al. (2025).

Our RCT directly evaluates the effectiveness of adding managers together with workers relative to adding workers alone. We randomly assign 112 rural government-run schools to one of three experimental arms: (1) a pure control group where schools continue to operate according to their status quo, without CEVs or a Fellow, (2) a government-implemented arm that mobilizes CEVs to provide out-of-school remedial education, or (3) an NGO-implemented arm that mobilizes both Fellows and CEVs. We assess student literacy and numeracy,

[1] Both literatures are extensive. See, for example, Schoellman (2011); Fujimoto et al. (2023) for its importance to macrodevelopment, and Bassi et al. (2020); Bold et al. (2017) for the impacts on student outcomes. The importance has been recognized for some time, with Hanushek and Woessmann (2008) offering a review of the older literature, and Aghion et al. (2025) offering a recent review.

measure the time use of head teachers and classroom teachers, and collect school-level characteristics, such as student enrollment and the number of classroom teachers, both at baseline and endline. The baseline took place at the beginning of the school year in which the intervention was introduced. The endline survey took place approximately nine months later, at the end of the same school year. Importantly, we assess students at endline regardless of their enrollment status.

Empirically, we find that adding a new level of management together with new workers significantly increases student numeracy, with the strongest impacts for those coming from low-asset households and those who recently re-enrolled. The program also leads head teachers to re-allocate their time away from certain non-managerial activities. In schools that receive a new manager (the Fellow) as well as additional workers (the CEVs), student numeracy increases by 0.33 standard deviations relative to the control group. In schools that receive workers without a manager (CEVs only), student numeracy increases by about half as much, 0.18 standard deviations. Moreover, in both treatment groups, learning gains are concentrated among students who come from households with fewer assets and among students who recently enrolled in school (i.e., were previously unenrolled), both of which are lower-performing groups absent treatment. In sum, the program is highly beneficial to student outcomes, especially for groups that exhibit lower learning levels in the status quo.

Further, we find substantial reallocation of time across tasks. CEVs alone enable teachers to focus significantly less on truancy and more on administration. Adding the Fellows as a new level of management leads to time re-allocation by head teachers, while adding CEVs alone does not. When a school receives Fellows, head teachers spend less time on individualized instruction and less time on truancy, tasks that may not be the highest value use of an administrator's time.

Armed with these experimental results, we develop a task-based production function for schools in which managers and workers of different types vary in their comparative advantage across tasks. In the model, the quantity of workers of different types is exogenous, and potentially suboptimal given costs, but workers and managers choose how to allocate their time across tasks to optimize student learning. We allow for the production function to be nonhomothetic in tasks: the optimal allocation of tasks changes with not only the input mix but also the the average learning level in the school.

To take the model to data, we use the first principal component across our four assessment scores as our scalar measure of output. Our approach uses the experimental variation in multiple ways. We estimate key parameters using our exogenous variation in workers as instruments for task increases and calibrate others to match treatment effects in the data. Combined with the experimental variation, the task-based structure of the model and the information on how workers sort identifies the overall production function that can be used to assess counterfactual staffing allocations. The estimated production function for learning is indeed nonhomothetic in tasks: individual instruction becomes more important as students achieve higher learning levels. CEVs and Fellows have a comparative advantage in truancy relative to teachers and principals, while Fellows have a comparative advantage in enrollment and classroom instruction relative to CEVs. Despite workers' ability to reallocate to new tasks, production exhibits diminishing returns to all resources as staffing increases.

We use the model, together with knowledge of the costs of personnel, to conduct counterfactual simulations that assess budget-neutral free lunches of two types. Both exercises are designed to utilize only modest extrapolations in staffing changes. In the first, we take the NGO's budget of the full intervention as given, and reallocate resources between CEVs and Fellows. We find that the optimal allocation would reallocate budget away from Fellows and toward CEVs, and the overall gains of this program are large, 36% of a standard deviation for the median program. Although budgets are the same across schools, the gains are even larger in lower performing schools. However, although the optimal allocation differs significantly from the current program, the gains of this optimal allocation relative to the current BT program are small, just 4% of a standard deviation. Hence, the current BT program approaches an optimal use of their budget.

The second counterfactual takes the *schools'* (rather than the NGO's) staffing budget as given but allows for modest (no greater than 10 percent) reductions in the budget spent on teachers and headteachers in order to reallocate toward CEVs and Fellows. These budget-neutral changes lead to gains of roughly one-quarter of a standard deviation in overall performance, with the lowest performing schools again gaining the most. This allocation is based on the current budget (without BT resources). Hence, the BT program is effective, but comparable gains could be had simply from reallocating existing staffing budget towards the resources that the NGO brings. There is indeed a free lunch.

Our paper contributes to literatures on school quality and organizational structure in low-income countries by showing how the the reallocation of staffing resources can lead to a free lunch for schools. There exists a large literature on the importance of school quality for development, including a micro literature assessing ways of improving school quality by increasing teacher quality (e.g., Bold et al., 2017), pay (e.g., De Ree et al., 2018), or effort (e.g., by reducing absenteeism as in Duflo et al., 2012), and adding supplemental inputs like books (e.g., Das et al., 2013), laptops (e.g., Beuermann et al., 2015), non-teaching staff (e.g., Ganimian et al., 2024), or early childhood interventions (e.g., Bassi et al., 2020). Some of these interventions increase student performance but conceivably also costs. We complement this literature by showing how the allocation of given resources can lead to a free lunch for schools, by reallocating them to their most effective use, which depends on the overall level of performance of the students. Our NGO partner's program of Fellows with CEVs is a practical example.

We also complement existing work on the staffing structure of organizations in low-income countries. Bloom et al. (2013) show that the quality of management drives firm performance in developing countries. They show that manufacturing firms that receive managerial consulting services increase productivity by 17%. Alfaro-Urena et al. (2022) show that domestic companies benefit from becoming suppliers to multinational corporations in part due to positive spillovers on managerial quality. Recent work by Bassi et al. (2025) shows that firms in low-income countries lack task specialization, reducing productivity, and Hjort et al. (2025) show that the relatively high cost of management deters structural transformation in low-income countries.

Our contribution is experimental evidence that empirically quantifies the benefit of adding workers with a manager relative to workers alone and showing how to use the experimental variation to estimate a task-based production function.

Much less work has empirically evaluated the impact of organizational changes in *schools* in low-income countries. Bloom et al. (2015), Kraft et al. (2016), and Dhuey and Smith (2018) show that the quality of the school principal matters for student performance in mainly high- and some middle-income countries. Lemos et al. (2024) document low management quality in schools in low-income countries and a positive correlation between management quality and school productivity, underscoring the need for work that quantifies the impact of managerial interventions in schools. Muralidharan and Sundararaman (2013); Duflo et al. (2015); Nyqvist and Guariso (2021); Singh et al. (2024) show learning gains of complementing traditional teaching teams with community members who may have lower formal qualifications than conventional government-hired teachers. We use an experimental evaluation to contribute new evidence on the impact of providing supplementary instruction via community-recruited staff alone relative to a setting where community-recruited staff work under a manager, and we show the impact of these changes on the time allocation of existing teachers and administrators.

The paper proceeds as follows. We describe the context in which the experiment takes place, the experimental design and data collection, the empirical results, the model estimation and calibration, and the results of counterfactual policy simulations.

2 Context

We partner with Building Tomorrow, a Ugandan-American non-governmental organization (NGO) working in the education sector, to study the organizational structure of schools in rural Uganda. Seeking solutions to achieve its goal of universal literacy and numeracy in Uganda, Building Tomorrow developed a community-powered model that adds a new layer of management to rural, government-run schools. The model consists of Fellows and Community Education Volunteers (CEVs).

Fellows. The Building Tomorrow model recruits recent Ugandan university graduates as Fellows, trains them on pedagogy and problem solving, and then deploys them to rural communities, where they work in four particular schools for two years. These Fellows come from various academic backgrounds including engineering, education, economics, hospitality, accounting and business administration. BT pays Fellows a salary and provides them with housing, a motorbike (for transportation to and from the schools where they work), a cell phone, and internet access during their two years of service. BT Fellows support classroom teachers and head teachers (akin to school principals) in their four assigned schools through engaging in daily tasks that vary between classroom instruction, lesson preparation, engagement with families and the community, one-on-one tutoring, administration, and implementing innovative pedagogical approaches.

Community Education Volunteers (CEVs). Fellows also recruit and train Community Education

Volunteers (CEVs). The CEVs, aged 17 to 70, are drawn from diverse backgrounds: they include parents, farmers, business owners, village leaders, and religious leaders. CEVs vary in the extent to which they have completed formal schooling but are selected for their commitment to enhancing local education. BT Fellows recruit eight CEVs per school. Once selected, CEVs receive training on mobilizing parents and guardians to enroll their children in both school and supplementary out-of-school instruction.

Both Fellows and CEVs supplement classroom instruction with out-of-school tutorials. The tutorials adapt the principles of Teaching at the Right Level (TaRL) for the rural Ugandan context.[2] The TaRL approach begins with assessments to identify each student's learning capabilities, enabling instructors to tailor their teaching accordingly (TaRL, 2025). This is a departure from one-size-fits-all pedagogy, and various studies have validated the effectiveness of using TaRL approaches to increase learning (e.g., Banerjee et al., 2007, 2016; Nyqvist and Guariso, 2021).

In 2023, Building Tomorrow served 1,500 primary schools in 38 districts across Uganda. Its Fellows and CEVs worked with about 180,000 students and re-enrolled 37,000 out-of-school children back in school (BT, 2025).

3 Empirical Methods: Experiment and Data Collection

We implement an RCT to evaluate the impact of two interventions that shock the labor structure of schools by adding staff of varying managerial levels. The first intervention is Building Tomorrow's full suite of programming (consisting of Fellows and CEVs who collectively provide supplementary TaRL-based instruction and conduct enrollment activities; here forward abbreviated as the "Fellows+" intervention). The second intervention is a government-coordinated program consisting of CEVs only, who provide supplementary TaRL-based instruction and conduct enrollment activities *without* the support of a Fellow (here forward abbreviated as "CEVs Only"). We assess the impact of each intervention on student enrollment and educational outcomes, including literacy and numeracy. We also conduct time use surveys to measure how different types of workers within schools (teachers and administrators) allocate their time across tasks.

3.1 Experimental Design

The experiment took place during the 2023 school year, in 112 rural, government-run primary schools across western, central and eastern Uganda.[3] We randomly assign each school to one of three arms (the control group, the Fellows+ arm, or the CEVs Only arm), and we stratify random assignment on geographic district (of which there are 18 in our sample). Figure 1 depicts the experimental design.

The Fellows+ intervention is the full BT program, inclusive of Fellows, CEVs, supplementary instruc-

[2]For example, BT staff add a social-emotional education component to address learning challenges related to Uganda's extended school closures for COVID-19. BT refers to its adaptation of TaRL programming as Roots to Rise (R2R) camps.

[3]In Uganda, the 2023 school year took place from March - December. School start was delayed from its typical starting date in February due to an Ebola outbreak.

tion, and enrollment activities. Fellows provide ongoing support to teachers and head teachers and lead the recruitment and training of CEVs, implementation of supplementary TaRL-inspired instruction, and student enrollment initiatives. This intervention includes all of the typical Building Tomorrow programming described in Section 2.

The CEVs Only intervention limits the additional layer of support to a government-coordinated program that recruited and trained CEVs to focus on the enrollment of out-of-school children and provide supplementary instruction. At the start of the academic year, local government officials, with support from some BT staff, recruited CEVs. Government officials, with support from BT, then provided a week of training in re-enrollment strategies and the provision of remedial education.

Both interventions may alleviate resource constraints by allocating additional labor hours to schools, at no cost to the school itself. In the first intervention, highly educated Fellows provide bespoke consulting services to head teachers and classroom teachers, akin to the consulting services provided to firms in Bloom et al. (2013), while the second intervention is a much less expensive program that primarily allocates more labor to schools. We evaluate the effectiveness of each program, assessing the extent to which an additional layer of highly trained management (in the Fellows+ intervention) impacts school functioning relative to additional labor alone (the CEVs Only intervention).

3.2 Data Collection

We documented enrollment status and administered a numeracy and literacy assessment to grade 4 (or "P4") students across all of 112 schools, at the beginning and the end of the 2023 academic year. Each student in the sample participated in two literacy assessments: one in English and another in their respective local language (Luganda, Lusoga, Runyoro, or Runyankole), corresponding to the primary language of instruction at their school. We administered the student survey and learning assessment for up to 20 P4 students per school.

Sample selection. In the Ugandan education system, the fourth year of primary school, P4, represents a critical learning year as students transition from lower primary school grades 1-3 to the upper primary school grades 5-7. Thus, we focus our evaluation on P4 students. We assessed all available students in schools with 20 or fewer P4 students at baseline. For schools with more than 20 P4 students, we randomly selected 20, ensuring that our sample represented the P4 population across the 112 schools. This approach yielded a baseline sample of 2,045 students, averaging just over 18 P4 students per school, indicating that many schools had fewer than 20 P4 students.

Sample retention. At the endline, we successfully located 1,938 of the initial 2,045 sampled students, a retention rate of 95%. We used similar survey tools at baseline and endline for students, grade four teachers, and head teachers. We varied the exact learning assessment from baseline to endline to limit the chance that the grade 4 teachers learn about the content of the assessment and "teach to the test." At the endline, we attempted to track and conduct the learning assessment for all students interviewed at baseline, regardless of their current enrollment status, i.e., the 95% retention rate is inclusive of both those who are enrolled and unenrolled at endline.

Measuring student learning. In order to evaluate the effectiveness of the two interventions on learning, the experiment employed the UWEZO assessment for all P4 students. Since 2009, the UWEZO test has been widely administered in Kenya, Tanzania, and Uganda on an annual basis to assess learning outcomes (GEM, 2014). UWEZO comprises a literacy and numeracy test. In Uganda, the literacy tests are in English and in seven local languages (Ateso, Lebacholi, Leblango, Luganda, Lusoga, Runyankore-Rukiga and Runyoro-Rutooro). The numeracy tests consist of an arithmetic evaluation and an ethnomath evaluation.

The UWEZO tests are dynamic, presenting the student with progressively more difficult questions if they answer a sufficient number of questions correctly on the preceding section. The order of the literacy test progresses from the identification of letters (lowest competency), to the identification of words, reading of paragraphs, and story comprehension (highest competency), respectively. The maximum score in English and the local language is four. In numeracy, the arithmetic evaluation progresses from counting (lowest level), to number recognition, and then to the operations of addition, subtraction, multiplication and division (highest level), respectively. Hence, the minimum score of zero (the student cannot count) and the maximum score is six (the student can do division). The ethnomath assessment presents students with two mathematical "word" problems, such that the maximum score is two (if the student answers both questions correctly).

Measuring time use and other school-level characteristics. We document the time use of teachers, head teachers, Fellows, and CEVs.[4] We ask each staff member which of 16 tasks they face in a typical semester and then follow up with a question about how many hours per week the staff member typically spends on each of the selected tasks. We aggregate the tasks into six major categories: Administrative, Enrollment, Individualized Instruction, Classroom Instruction, Preparation and Truancy. Each classroom has one teacher who is responsible for instruction in that room, and each school has one head teacher administering the teachers.[5] Finally, we collect other school-level characteristics in the head teacher and classroom teacher interviews, including total school enrollment, total teachers per school, utility (e.g., water and electricity) access at the school, and teachers' educational attainment.

[4]In our sample of schools, teachers and head teachers compose the complete set of workers in schools, absent our intervention. External volunteers and other forms of non-teacher support staff were not found in our sample. Hence, we are confident that we are capturing the full set of staff in schools at baseline.

[5]Head teachers play many roles. In addition to handling administration and interacting with the teachers and the community, head teachers often directly teach when, for example, a teacher is truant. All of this is captured in our time use data.

4 Empirical Results

We estimate treatment effects from our intervention by comparing changes in outcomes at the school or student level across randomized treatment groups. We estimate the following primary specification:

$$y_{ijt} = \alpha_i + \sum_g \beta_g \mathbb{1}[\text{Treatment Group}_{jt} = g] + \gamma X_{ijt} + \varepsilon_{ijt}, \tag{4.1}$$

where i is the individual, j is the school, t is the wave (baseline or endline), $g \in \{\text{Fellows+, CEVs Only}\}$, and X_{ijt} are controls.[6] For inference, we clusters residuals at the district level (there are 18 districts).

4.1 Student Outcomes

Our primary outcome of interest is how students performed on the UWEZO assessments of numeracy and literacy. For each assessment type (numeracy and literacy) we consider four types of outcomes. First, we estimate the effects on the number of completed sections, which we express in standard deviation units.[7] Second, we estimate whether or not students reached threshold proficiency levels differentially in the treatment groups compared to control.[8] Third, we estimate the effect on whether or not the student *progressed* by completing strictly more sections at endline than at baseline. Finally, we estimate the treatment effect on *learning loss* by considering whether or not the student finished strictly *fewer* sections at endline than at baseline.

In Table 1 we report treatment effects on the number of sections completed. The first four columns simply correspond to the four sections of the assessment. Numeracy is the equally-weighted sum of math and ethnomath, and literacy is the equally-weighted sum of English and local language. All six of these measures are divided by their baseline standard deviation so that the treatment effects are interpreted in standard deviation units.

Both the Fellows+ and CEVs Only arms increase student numeracy. The Fellows+ arm increases math scores ($p < 0.10$) by 0.15 standard deviations (sd) above the gain realized by the control group over the course of a single school year and increases ethnomath scores ($p < 0.05$) by 0.36 sd above the control.[9] In ethnomath and overall numeracy, the Fellows+ program arm raises student achievement by roughly twice as much as the CEVs Only program. Throughout this section, this is consistent across our outcome variables.

Since we conducted our baseline assessments at the beginning of the academic year and our endline assessments at the end of the year, we can use these estimates to ask how the average rate of progress was affected by the intervention. For instance, in Math the average student's assessment score in the control group was 0.37 standard deviations greater at endline than it was at baseline. Hence, the point estimate of

[6] Throughout, we include district-wave fixed effects and enumerator fixed effects.
[7] Please refer back to Section 3.2 for a complete description of the assessments.
[8] For numeracy, we consider whether or not the student can do addition and subtract. For literacy, we consider whether or not the student can read simple words. In both categories, this is the median endline competency in the control group.
[9] The math test poses arithmetic problems, while the ethnomath test poses mathematical "word" problems.

0.15 standard deviations in the Fellows+ program arm implies that the average student in this group made 40% more progress in that academic year than the average student in the control group.[10]

We find stronger effects from both interventions on Ethnomath, and on Numeracy (which is the sum of the two). These two categories showed no progress in the control group (in fact, the negative effect on Ethnomath means a meaningful number of students did worse at the end of the year than the beginning in that category). Hence, in these categories we can say that the interventions cause progress in learning categories that have no progress in the control group.

We cannot detect a statistically significant effect of either intervention on English scores, local langauge scores, or overall literacy. For both interventions, the treatment effects are positive but imprecise, leaving open the possibility that the current study lacks power to detect small positive effects on those outcomes. This result is consistent with others in the education literature that indicate greater difficulty in improving literacy than numeracy (e.g., Rich, 2013; Chetty et al., 2014).

Next, instead of using scores on the assessments as our outcome variable, we use an indicator variable equal to one if the student has reached a threshold proficiency level and zero otherwise. These results are reported in Table 2. First we look at whether or not students complete the section that demonstrates they can add and subtract, and we find positive and significant treatment effects from both treatment arms.[11] We find positive but not significant treatment effects on division, the highest proficiency in our assessment. Likewise, we find positive, not significant effects in the demonstrated ability to read words. We do find a positive and significant effect (on the government-coordinated program of CEVs arm on paragraph reading comprehension, the highest proficiency in the English assessment).

In terms of measured progress, we note that these results imply that in the control group 14% more students can add and subtract at endline than could at baseline. The treatment effect in the Fellows+ program of 11% thus implies 77% more progress than in the control group. In the government-coordinated program of CEVs group, a treatment effect of 6% implies 42% more progress than in the control group.

Finally, we take as our outcome variable whether or not a student improved or regressed in their assessment between the baseline and endline. These results are reported in Table 3. In Column 1, our outcome variable is an indicator function that takes a value of one if the student's assessed Numeracy is strictly greater when measured at endline than it was at baseline. Here we find that students in the Fellows+ arm were significantly more likely to improve in their numeracy assessment than students in the control group. The effect in the government-coordinated program of CEVs arm was positive but not significant.

The second column shows that the probability that students did strictly worse on the endline assessment than the baseline. We find that this type of measured learning loss is alarmingly common in the control group: 44% of students did strictly worse at endline than at baseline. We find that this regression in numeracy

[10] This calculation is simply 0.15/0.37 = 40%. Even without the regression estimates, we can simply compute the mean increase in test scores in the treatment and control groups. We find the are 0.369 in the control group and 0.487 in the Fellows group, implying 32% more progress.

[11] We find that this is the median math proficiency level in the endline of the control group, so this outcome shows whether or not students are achieving progress typical of their grade level in this context. Likewise, we use the ability to read words as the literacy threshold for the same reason.

is significantly mitigated in both treatment arms. Here, negative treatment effects mean that significantly *fewer* students had learning loss in numeracy in the treatment groups.

The point estimates on Literacy mirror the results in Numeracy, but the effects are smaller and not statistically significant. Once again, this result is consistent with others in the education literature (e.g., Rich, 2013; Chetty et al., 2014).

To summarize, both treatment arms improve student numeracy. Effects are consistently larger in response to the Fellows+ program relative to CEVs Only. We see stronger effects on mitigating learning losses than on increasing learning gains, suggesting that the intervention prevented students from falling behind as opposed to increasing the fraction of high-perforrmers. Moreover, the largest treatment effects occur for those who have low assets and are newly enrolled. In effect, the Fellows+ program mitigated learning loss for those learners most at-risk of falling behind.

4.2 Heterogeneity Analysis

We now test for differential treatment effects based on baseline student characteristics. Here our specification is:

$$y_{ijt} = \alpha_i + \nu_t C_i + \sum_g (\beta_g \times \delta_g C_i) \mathbb{1}[\text{Treatment Group}_{jt} = g] + \gamma X_{ijt} + \varepsilon_{ijt}, \tag{4.2}$$

where C_i is an indicator function for a binary baseline characteristic of individual i. Here we include the characteristic interacted with a time fixed effect to allow for the possibility that the characteristic predicts growth in the outcome variable.[12] Differential treatment effects here are given by δ_g. Note that the effects reported throughout this subsection are additive, so the total effect of the intervention on a person with characteristic C_i is $\beta_g + \delta_g$.

We begin by estimating the differential impact of the treatments on female students, which we report in Table 4. Here we list the four outcomes that had significant treatment effects in the previous subsection. Here we find no statistically significant evidence of differential treatment effects for female students. However, we do note that the effects on Numeracy and threshold proficiency in addition and subtraction have quantitatively meaningful point estimates, but with wide confidence intervals.

Next we consider a measure of assets that we use to proxy the socioeconomic status of the households students live in. To measure this, our baseline survey includes discrete questions on whether or not the student has various assets at home.[13] Our measure of socioeconomic status is the number of the assets that they report owning. We call a student "high asset" if they are above the median by this measure.

Table 5 reports effects differentially by this measure. We find that students with below-median assets see broadly greater benefits from the interventions. In the government-coordinated program of CEVs arm

[12]Note that given the presence of the individual fixed effect if $t \in \{1, 2\}$, then of course ν_1 is not identified. We are only interested in ν_2, the differential growth rate in the outcome variable.

[13]The seven assets we ask about are: electricity, mobile phone, television, bicycle, motorcycle, refrigerator, and separate kitchen. By design, these are all things a young student would readily know in this context. We find that students own a mean of 3.3 of these assets with a standard deviation of 1.37.

there are positive and statistically significant differential effects on the ability of low asset students to achieve proficiency in addition and subtraction. Likewise, there are negative and statistically significant differential effects on the probability of regressing in numeracy. In the Fellows+ group, these effects are similar, but not statistically significant.

Finally, we test whether or not there are differential impacts for students that were not enrolled in the school in the prior school year. We call these students "newly enrolled." This is particularly important in the context of rural Uganda because there are high rates of inconsistent enrollment. In particular, schools in Uganda were closed for nearly two full school years during the COVID-19 pandemic, and students have been slow to reenroll. Students often fall behind, and do not feel they can progress after being out of school for so long. We are interested in testing if the interventions are particularly effective at helping students who may be behind because they are newly enrolled compared to those enrolled at the school in the previous year.

We report these effects in Table 6. We find positive effects on numeracy from the interventions (significant in the Fellows+ group; not significant from the CEVs Only group, though with similar point estimates). As anticipated, in the control group the newly enrolled students show lower progress compared to continuing students, with a significant negative difference. The magnitude of the positive differential effects are very quantitatively similar to this effect. As a result, these estimates suggest that not only do the continuing students get positive effects but that newly enrolled students completely close the progress gap between them.

Taken together, these results show that the gains from the intervention are concentrated in students whose baseline characteristics predict slower progress (low assets or newly enrolled). This is consistent with the interpretation of the results from the previous subsection, which suggested that the gains are strongest in outcomes related to improving the outcomes of lower performing students.

4.3 Task Composition

To better understand the mechanisms through which adding staff to schools affects student outcomes, we next turn to our data on time spent on tasks by staff in schools.

First, in Table 7 we show how teachers divide their time across arms. In the control group we can see that the two largest tasks by far for teachers are Classroom Instruction and Preparation, which together account for almost 80% of teacher time. In the treatment arms we find that they shift their time away from Truancy tasks and Enrollment (though not statistically significantly). We see they shift their time toward Administration and Individualized Instruction.

Second, Table 8 shows the task distribution of head teachers. In the control group, we can see that head teachers spend much of their time on Administration, as expected, but that they also spend nearly half of their time on Classroom Instruction and Preparation. We see that in both treatment arms they reallocate their time away from Truancy and Individualized Instruction, and toward Enrollment and Classroom Instruction.

To understand this reallocation of tasks by teachers and head teachers, we next look at how CEVs and

BT Fellows that are experimentally added to schools spend their time. This is summarized in Table 9. We can see that BT Fellows spend most of their time on Individualized Instruction, but also they spend significant time on Enrollment. We see that CEVs may also spend time on Individualized Instruction, but they change their time allocation depending on whether or not a BT Fellow is also present. In particular, CEVs spend significantly less time on Administration and Enrollment when a BT Fellow is present (though the difference is only statistically significant for Administration). Consistent with this time allocation, the CEVs Only program significantly reduces the proportion of students who are unenrolled at the endline relative to the control (see Table A.1 in Appendix A). The presence of a Fellow appears to shift the use of CEV time away from enrollment activities and toward instruction, which could be a change in CEV's comparative advantage in the presence of the Fellow or the fact that once enrollment activities are sufficiently accomplished, individualized instruction becomes a bigger priority.

Taken together, we see some clear patterns but also complex patterns. Given the treatments, teachers and head teachers reallocate their time away from Truancy, which BT Fellows and CEVs spend time on. This suggests that the effort of different workers may be substitutes in some tasks. However, in both treatment arms teachers spend more of their time on Individualized Instruction, but that is precisely the task that CEVs and BT Fellows spend most of their time on. This suggests that the relationship is more complicated than pure substitution alone.

More broadly, CEVs alone seem to impact teachers' time allocation, while the addition of Fellows impacts the the time allocation of headteachers. This may suggest that like headteachers, Fellows are managers, and like teachers, CEVs are workers, but again it doesn't seem to be a pattern of pure substitution: the impact of CEVs on teachers disappears when Fellows are present, and so the four types of staffing interact in more complicated ways.

To get a complete picture of the total time spent on each task by all workers in the school, we report the total hours spent on each task adding across all worker types. Here, we take the mean number of hours within each task by worker type by treatment arm level. We then divide the number of hours by the number of students served by that worker.[14] Then we add across workers to compute the total hours applied per student per term.

We find large differences across tasks in the total number of hours applied in the different treatment arms. Although we saw reallocation of time away from Truancy by head teachers and teachers, it is almost exactly offset by time spent on Truancy by CEVs and BT Fellows.

Among the other tasks, total time spent on each increases but by very different amounts. The category that increases the most in percentage terms in Individualized Instruction, which more than doubles in the CEVs Only arm and more than triples in the Fellows+ arm. Preparation increases by 6.3% in the CEVs Only arm and 26.2% in the Fellows+ arm. All other tasks change by amounts intermediate to those. These large

[14]For teachers, the number of students served is the number of students in the classroom. For CEVs and head teachers, we divide by the number of students in the school. For fellows, who are assigned to four schools simultaneously, we divide by their time by the number of students times four.

differences in task composition in response to changes in staff structure suggest rich patterns of substitution and prioritizing across tasks.

5 Estimating a Task-Based Schooling Production Function

Motivated by our empirical results, we now show how to use our experimental design and data to estimate the education production function. In the next section, we will then use the parameterized model to perform counterfactuals and make predictions about how education outcomes would change with modified worker structures within schools.

Suppose the output of school j is y_j, such as average assessment scores or the fraction of students meeting a given standard. The outcome depends on how much of each task $k \in K$ (such as exam grading, administration or lecturing) is performed within the school j, T_{jk}. Within each school each worker has a type $i \in I$ (such as teachers or head teachers) and exerts effort toward each task. The time allocated by a worker of type i at school j on the task k is l_{ijk}.

Workers of type i are endowed with E_i of effort, which they use to complete tasks. We do not assume that the marginal cost of effort on each task is constant, however. In particular, the amount of effort required to spend l_{ijk} hours on a task need not be linear. The amount of labor a worker is capable of expending across tasks is given by:

$$\forall(i,j), E_i \geq \left(\sum_k \theta_{ik} l_{ijk}^{\gamma}\right)^{1/\gamma} \tag{5.1}$$

where θ_{ik} is the productivity (or its reciprocal is the effort cost) of a worker of type i engaged in task k. The parameter γ governs the substitutability in cost of effort in performing different tasks. Marginal effort is given by:

$$\frac{\partial}{\partial l_{ijk}} \sum_{i'} \theta_{i'k} l_{i'jk}^{\gamma} = \gamma \theta_{ik} l_{ijk}^{\gamma-1} \tag{5.2}$$

so that if $\gamma > 1$ marginal effort is increasing – focusing on a single task requires more effort – whereas if $\gamma < 1$, it is decreasing in a task, so multi-tasking requires more effort. If $\gamma = 1$, marginal effort is constant.

Note that effort itself is unobservable, only hours are observable, and they will be endogenous to the tasks assigned. Because workers of type i are identical and equation(5.1) exhibits constant returns to scale, we can scale up workers, effort, and hours of type i linearly. The total amount of labor in school j in task k is therefore given by the sum of labor across all worker types:

$$\forall(j,k), T_{jk} = \sum_i l_{ijk} \tag{5.3}$$

Finally, we assume that the total output (e.g., test scores) depends on an aggregation of all these tasks. In particular, we want to allow for the possibility that aggregation is non-homothetic. This is important in this context, because we want to account for the likelihood that many tasks within schools are unlikely to scale proportionally with labor, such as classroom lecturing. Since we are particularly interested in how tasks *shift* with additional labor, we want to model this explicitly.

To allow for non-homotheticity in a way that will be consistent with a simple method of parameter identification, we assume that the school output y_j is defined implicitly by a non-homothetic aggregator of the form:

$$\forall j, \qquad 1 = \sum_{k \in K} Z_j \left(\frac{T_{jk}}{y_j^{\rho_k}} \right)^{\sigma} \tag{5.4}$$

where σ controls how substitutable tasks are with one another. The parameter Z_j is school-level productivity, accounting for unexplained differences in test scores across schools.

The parameters ρ_k play two roles. The first roles can be seen easily in the special case that $\rho_k = \bar{\rho}, \forall k$. In this case, $1/\rho$ would capture *the returns to scale* in translating task output to overall school performance. More generally, because the extent of increasing or decreasing returns to scale can be task-specific with $\rho_k \neq \rho_{k'}$, the second function is to yield an optimal task assignment that depends on the school's overall performance, y_j, and hence also the total quantity of labor in the school. That is, from any starting allocation of labor if all forms of labor doubled, this *non-homotheticity* would imply relative task reallocation. Tasks with higher values of ρ_k have lower returns to scale and so their relative productivity falls as staffing resources expand. In the case of low substitution, $\sigma < 1$, these tasks therefore receive a relatively larger share of labor as labor expands.[15]

Finally, note that there is an Inada condition on tasks – all tasks will be performed at some level – but not on workers types. Since workers only provide tasks, the marginal benefit of adding worker of a new type is bounded, and may or may not exceed its marginal cost. Moreover, equation (5.1) also exhibits no Inada condition in terms of driving the marginal cost of providing labor to zero. If $\gamma > 1$ workers will optimally diversify across tasks, but the marginal cost of working is always positive. Simply put, there is no hard-wired gain to expanding the variety of workers.

5.1 Time Allocation Problem

Then the problem of the school takes the number of workers of each type, the E_i terms, as given and chooses l_{ijk} to maximize y_j. The problem of school j is given by:

$$\max_{\{y_j, T_{jk}, l_{ijk}\}} y_j \tag{5.5}$$

subject to equations (5.1), (5.3) and (5.4).

[15] In consumer theory, if these were preferences we would think of ρ_k as determining *income effects*, where consumption would shift toward goods with higher ρ_k as income increases. This is conceptually similar, but the notion of "income" here would be about the total labor instead. Also, since this is production, we cannot simply normalize one ρ_k because it is linked to a cardinal level of output, rather than ordinal utility. Indeed, the scale of ρ_k captures the returns to scale.

The optimal allocation of time by a worker of type i on task k is given by the first-order conditions on time allocation:

$$l_{ijk}^{\gamma-1}\theta_{ik}\gamma\eta_i = \frac{T_{jk}^{\sigma-1} y_j^{-\sigma\rho_k}}{\sum_{k'} \rho_{k'} T_{jk'}^{\sigma} y_j^{-\sigma\rho_{k'}-1}} \tag{5.6}$$

where η_i is the Lagrange multiplier on equation (5.1). Understanding this first order condition can be greatly simplified by looking at l_{ijk} *relative to* labor by the same worker type in a reference task, which we will label task 0. Then:

$$\frac{l_{ijk}}{l_{ij0}} = \left(\frac{\theta_{ik}}{\theta_{i0}}\right)^{\frac{1}{\gamma-1}} \left(\frac{T_{jk}}{T_{j0}}\right)^{\frac{\sigma-1}{\gamma-1}} y_j^{-\sigma\frac{\rho_k-\rho_0}{\gamma-1}} \tag{5.7}$$

That is, this says that the optimal relative allocation of labor of given workers type in a given school across tasks is determined by their relative productivity in those tasks (θ_{ik}/θ_{i0}), the relative allocation of time in the school (T_{jk}/T_{j0}), and the level of output (y_j). Notice that this last term only appears because the model is non-homothetic. Indeed, a special case of the model is when $\forall k, \rho_k = 1$, in which case there is constant elasticity of substitution across tasks and the model is homothetic. As we can see here, in that case y_j disappears from this equation.

5.2 Identification of Model Parameters

We now map the model to our particular setting with four types of workers, i.e., teachers, ($i = 1$), head teachers ($i = 2$), CEVs ($i = 3$), and Fellows ($i = 4$). Given the above, we show that most of the parameters of the model can be recovered from a linear regression, for which we will use the treatments as instruments for identification. In particular, note that we can rewrite equation (5.7) as:

$$\log(l_{ijk}) - \log(l_{ij0}) = \frac{1}{\gamma-1}\log(\theta_{ik}/\theta_{i0}) + \frac{\sigma-1}{\gamma-1}\log(T_{jk}/T_{j0}) - \sigma\frac{\rho_k-\rho_0}{\gamma-1}\log(y_j) \tag{5.8}$$

where l_{ijk} is time spent by workers of type i in school j on task k, and in this setup differencing is with respect to a reference task (labeled 0). Here, T_{jk} is the total labor on task k at school j, and y_j is test scores at school j. The parameters σ (substitutability across tasks), γ (effort cost curvature), and $\forall k, \rho_k$ (task composition with level of y_j) are parameters to be estimated, and the productivity terms (θ_{ik}) are related to fixed effects.

In the model, we are interested in how the labor allocation of an individual worker type responds when the total amount of labor of that type in the school changes. Obviously, this is endogenous: whatever might cause a change in demand for labor of a particular type of labor in the school would likely affect the choices of each type and, therefore, the total labor allocated to that task in the school. Hence, we would expect an OLS estimate of this relationship to be positive: the allocation choice of worker types and the whole school move in the same direction.

Yet task shifting would imply the relationship is of the opposite sign. When more labor is exogenously added to the school that takes on relatively more of a task, each worker type responds by shifting away from those tasks and toward tasks for which they have a comparative advantage.

In addition, another challenge in treating equation (5.8) as an estimation equation is that the contemporaneous output of the school y_j is clearly endogenous to the labor choices. To address this, instead of using the endline school outputs directly, we proxy y_j (test scores) with a polynomial function of baseline test scores estimated in the control group, which we label $\tilde{y}_j$.

Now we estimate the equation:

$$\log(l_{ijk}/l_{ij0}) = \alpha_{ik} + \beta \log(T_{jk}/T_{j0}) + \sum_{k'} \delta_{k'} I\{k' = k\} \times \log(\tilde{y}_j) + \varepsilon_{ijk} \tag{5.9}$$

From this, we can then see that the point estimates here are related to the model parameters as follows:

$$\beta = \frac{\sigma - 1}{\gamma - 1} \tag{5.10}$$

$$\forall k \neq 0, \delta_k = \sigma \frac{\rho_k - \rho_0}{\gamma - 1} \tag{5.11}$$

If we take as given the value of γ and ρ_0, we can then solve for σ and the other ρ_k parameters as:

$$\sigma = 1 + \beta(\gamma - 1) \tag{5.12}$$

$$\forall k \neq 0, \rho_k = \rho_0 + \frac{\delta_k}{\frac{1}{\gamma - 1} + \beta} \tag{5.13}$$

We are then left with the question of how to identify γ and ρ_0. To do this, we match two moments from the data: the treatment effect of the Fellows intervention on test scores, and the treatment effect of the Fellows intervention on the Herfindahl index across tasks for teachers. Both moments are affected by both γ and ρ_0, but the magnitude of treatment effects on test scores is most closely tied to ρ_0, since that determines how much additional task time affects test scores. Conversely, the Herfindahl index across tasks is most closely tied to γ, because it controls how easily a given worker type can substitute between tasks.

5.3 Estimating the Model

The overall idea behind the estimation is to link the data we have on task allocation by different inputs, test score outcomes, and experimental variation in the two from staffing differences to estimate the key structural parameters which will allow for extrapolation beyond the experimental results.

We begin by estimating equation (5.9), where we set the reference task as Preparation.[16] For all inference we cluster standard errors at the district level.

[16] Preparation is chosen because it is the task with the smallest cross-worker type $\mathrm{Var}(\log(l_{ijk}))$ in the data. That is, it is the task on which time spent is the most similar across worker types.

As detailed in prior sections, we gather data on several different tests and therefore could measure school-level output in several ways. We choose the first principle component of the available test scores as a single index for output.[17]

We consider four specifications, interacting two different production functions with two different estimation strategies. The "Non-Homothetic" production function is as described above, while the "Homothetic" specification drops $\Delta \log(\tilde{y}_j)$ from the right hand side, which corresponds to the model in which $\forall k, \rho_k = \rho$. In the "OLS" estimation, we ignore the identification problem and simply estimate equation (5.9) directly. In the "IV" estimation, we use the two different treatment indicators as instruments for the potentially endogenous regressor $\Delta \log(T_{jk}/T_{j0})$. Intuitively, the nature of the intervention is to add labor to schools, which is likely to increase the amount of total labor in each task within treated schools. Moreover, these effects are likely different depending on the type of labor added (hence, two different instruments corresponding to the two treatment arms). This procedure is detailed later in this section.

In Table 11, comparing the IV and OLS columns we can see the importance of our identification strategy. The point estimate on the change in total task time is positive in the OLS but negative after using the instruments. This is consistent with the explanation provided before: if schools differ in their demand for particular tasks, then schools with greater demand for a particular task have more of each worker type spending time on that task and therefore more total time spent on the task. The IV interpretation is when more labor is exogenously added to the school, workers shift *away* from the tasks receiving the largest inflows of labor, consistent with the negative point estimate. The first stage F-statistics are approximately 23 in both cases, giving us some confidence that these results are not driven by weak instruments.

Next, we can evaluate whether or not the results in the data are consistent with non-homotheticities by looking at the point estimates on the interactions between changes in test scores and each task. The homothetic model predicts that all these interaction terms should be equal to zero, so we can evaluate the null hypothesis that the model is homothetic with the joint test that all the interaction terms are jointly equal to zero. Our Wald Chi-squared hypothesis test rejects the null hypothesis with a p-value of 0.006. This gives evidence that the data is consistent with the non-homothetic model.

Among the interaction terms in the Non-Homothetic IV column, we can see that only two are statistically significant at the 10% level. In both cases they are positive, which the model interprets as tasks that would the school would optimally assign a greater fraction of labor when test scores increase.

5.4 Calibration of Remaining Parameters

To solve the full model, we need to recover a few more parameters that are not directly implied by the estimation strategy implemented in the previous subsection. In particular, we still need to find γ (the shape

[17] An alternative would be to follow the broader education literature and measure output the fraction of students that demonstrate proficiency in addition and subtraction in the endline survey round. However, this would be limited on two fronts. First, it would not fully utilize the breadth of our data, focusing on only one learning outcome. Second, it would completely ignore any gains beyond proficiency.

of the effort cost across tasks), ρ_0 (the curvature parameter on the reference task), and the matrix of θ_{ik} (productivity at each task k for each worker type i).

First, to get the matrix of productivity terms θ_{ik} we match the mean time spent by each worker type i on each task k in the Fellows treatment arm.[18] To do this, we use the optimality conditions of the assignment problem to solve for θ_{ik} as a function of l_{ijk}. To do this, we start by normalizing the total effort budget for all workers types to one, $\forall i, E_i = 1$. Then we can rewrite equation (5.6) as:

$$\theta_{ik} l_{ijk}^{\gamma} = \frac{l_{ijk}}{\gamma \eta_i} \frac{T_{jk}^{\sigma-1} y_j^{-\sigma\rho_k}}{\sum_{k'} \rho_{k'} T_{jk'}^{\sigma} y_j^{-\sigma\rho_{k'}-1}} \tag{5.14}$$

Then we substitute this into equation (5.1) with the $E_i = 1$ normalization in order to solve for the Lagrange multiplier η_i:

$$1 = \sum_k \theta_{ik} l_{ijk}^{\gamma} = \frac{1}{\gamma \eta_i} \frac{\sum_k l_{ijk} T_{jk}^{\sigma-1} y_j^{-\sigma\rho_k}}{\sum_{k'} \rho_{k'} T_{jk'}^{\sigma} y_j^{-\sigma\rho_{k'}-1}} \implies \eta_i = \frac{\sum_k l_{ijk} T_{jk}^{\sigma-1} y_j^{-\sigma\rho_k}}{\gamma \sum_{k'} \rho_{k'} T_{jk'}^{\sigma} y_j^{-\sigma\rho_{k'}-1}} \tag{5.15}$$

Substituting this back into equation (5.6) and rearranging implies:

$$\theta_{ik} = \frac{l_{ijk}^{1-\gamma} T_{jk}^{\sigma-1} y_j^{-\sigma\rho_k}}{\sum_k' l_{ijk'} T_{jk'}^{\sigma-1} y_j^{-\sigma\rho_{k'}}} \tag{5.16}$$

Then conditional on the parameters σ, γ and all the ρ_k terms, we can calculate the matrix of θ_{ik} terms as a function of observable labor allocations and test scores. Moreover, with the relationships shown in equations (5.12) and (5.13), conditional on γ and ρ_0 the point estimates from the IV allow us to identify σ and all other ρ_k terms. Therefore, we can now fully parameterize the model if we know γ and ρ_0.[19]

To get γ and ρ_0 we calibrate the model to exactly match two treatment effects from the Fellows arm. We interpret the intervention in the context of the model as changing the E_i terms. In the control group, we assume $E_1 = E_2 = 1$ and $E_3 = E_4 = 0$. The Fellows arm sets $E_1 = E_2 = E_3 = E_4$ to one. The calibration strategy is to compute the model for the control group and the Fellows arm and compute two moments: the treatment effect on test scores, and the amount of time teachers spend on classroom teaching. The treatment effect on test scores we seek to match is 0.33, while the treatment effect on time spent on teaching 0.13. γ capture the marginal gains of additional resources, and ρ captures how the overall scale of non-homotheticities of different tasks.

Applying these procedures, we calibrate the model and find parameter values in the non-homothetic and homothetic versions of the model. These are given in Table 12. In addition, in Appendix B we list the values of the worker type-task cost terms ($\log(\theta_{ik})$).

[18] We choose the Fellows arm because it is the only arm in which we observe the time allocation of all four types of workers: teachers, head teachers, CEVs and Fellows.

[19] Though not emphasized here, there is also the aggregate productivity parameter Z_j, which scales labor allocations to test scores. Since we match the time allocation exactly and observe test scores, this can be computed in closed form from equation (5.4): $Z_j = 1/\sum_k (T_{jk}/y_j^{\rho_k})^{\sigma}$.

5.5 Comparative Statics

Figure 2 demonstrate the comparative statics in the estimated model, showing the increases in average quality from adding CEVs and Fellows. We measure the average quality in standard deviations (of the first principal component of test scores) and each of the inputs in units of their treated level (i.e., 1 is the treatment level). In model nomenclature, denote $y_j^*(E_1 = 1, E_2 = 1, E_3, E_4)$ as the optimized output for school of type j given the stated resources. We will vary E_3 (CEVs) and E_4 (Fellows). We present three separate comparative statics exercises. For each comparative static, we plot the additional score increase in three different quality schools: the 10th percentile, median, and 90th percentile.

The first exercise corresponds to a model-implied intensive margin treatment within our CEV-only treatment arm, where the intensive margin is extrapolated based on the estimated task-productivities from the additional information about task allocation of CEVs (and others). We plot $y_j^*(1, 1, E_3, 0) - y_j^*(1, 1, 0, 0)$. The exercise increases the CEVs in a school that has the average level of teachers and head teachers (per student) and no Fellows. The results from this experiment are the three upper most (i.e., solid lines) curves in the the figure. This treatment has the highest impact on its own, but it varies by the type of school. In a school with lower baseline test scores the treatment impact is strongest, and the difference between the 10th percentile school and the 90th percentile school can be sizable. This shows the importance of the nonhomothetic production function. Nevertheless, the diminishing returns to additional CEVs is evident in each of the three schools.

The second exercise and third exercises are linked to intensive margins of the "Fellows +" treatment arm, both where we separately vary the intensive margin of CEVs and Fellows. The second exercise is therefore to increase the CEVs in a school that has the average level of teachers and head teachers (per student) *and* one Fellow, i.e., $y_j^*(1, 1, E_3, 1) - y_j^*(1, 1, 0, 1)$. These results are plotted in the lower most (i.e., double lines) curves of the figure. Here, the CEVs again exhibit diminishing returns and again have a sizable impact, but it is markedly less than in the case of no Fellows. Moreover, the difference in impact in high- and low-quality schools is much smaller than when CEVs are added without Fellows.

The third exercise is to increase the intensive margin of Fellows in a school that has the average level of teachers and head teachers (per student) *and* one CEV, i.e., $y_j^*(1, 1, 1, E_4) - y_j^*(1, 1, 1, 0)$. The results are plotted as the dashed line, and they fall in the middle of the other two comparative static exercises. The results again show diminish returns, with gains that are greater than CEVs with a Fellow but less than CEVS without a Fellow, and the low-performing school benefits more than the high-performing school, thought the difference is not as extreme as in the case with CEVs but no Fellow. Of course, these comparative static exercises ignore the costs of the inputs, which we turn to now.

6 Counterfactual Policies

We now use the fully parameterized model to conduct counterfactuals where we exogenously change the labor structure of the school and study the effects on test scores in search of free lunches. To look for free lunches, the counterfactuals are therefore implemented as budget neutral changes in the total amount of E_i available to the school subject to particular budget.

We consider two families of counterfactuals. In the first set, we consider changes to the NGO's treatment that are budget neutral from the NGO's perspective. In the second set of counterfactuals, we consider changes to overall staffing that are budget neutral in terms of the overall cost of staffing.

Both exercises require data on the annual cost of labor of each type to evaluate tradeoffs between worker types. The data for CEVs and Fellows come from the program cost of our NGO partner as implemented, and without any evaluation costs.The data on teachers and head teachers comes from average salaries reported by the Ministry of Public Service. We summarize this information in Table 13.

In absolute terms, Fellows are the most expensive, followed by head teachers, and teachers, with CEVS being by far the least expensive. However, the second row indicates that there are many more teachers per school (7.13) than CEVs (4), head teachers (1), and Fellows (0.25). The second row becomes our choice of units: i.e., one model unit of teachers is 7.13 actual teachers, whereas one model unit of Fellows is 0.25 actual. The bottom row normalizes all staffing expenses by average teacher expenses. We can see that teachers are the most expensive, followed by head teachers and Fellows, with CEVs by the far the least expensive. We use this budget information to consider staff changes that keep the total budget fixed but change the composition of labor in the school.

6.1 Budget-Neutral NGO Staffing

Given our normalization, current Building Tomorrow program adds on CEV and one Fellow to each school, which costs (0.05+0.21=)0.26 teachers. We consider reallocating those resources optimally across CEVs and Fellows, keeping the total budget in each school unchanged. We allow the allocation to vary across school quality, however.

The results are presented in Table 14. Given the very low costs of the CEVs, the first two rows show thatthe model implies that the optimal solution would be to roughly quadruple the size of the CEVs program, and cut the Fellows to roughly 30% of its cost. The overall gains of the program overall would be large, as shown in the third row, increasing overall schooling test scores by 40% of a standard deviation in the 10th percentile school, and one-third of a standard deviation in the 90th percentile school with a median gain of 36% of a standard deviation. The implied allocation goes far beyond the experimental bounds of our data, but the exercise actually is useful in telling us that these extrapolated gains are actually quite small. Even if one were to palate these extrapolations, one would only gain 0.03-0.04 standard deviations by reallocating. The BT program which is both effective and cost-effective is close to optimal given the financial resources it uses.

6.2 Budget-Neutral School Staffing

We now ask whether there are gains to reallocating school resources across types of staffing. Given our normalization, the current school budget (without Building Tomorrow's resources) is one teacher and 0.22 head teachers, for a total budget of 1.22 teachers. We consider reallocating those resources optimally across CEVs and Fellows, keeping the total budget in each school unchanged. We again allow the allocation to vary by school quality. However, we are interested in evaluating potentially palatable reallocations that are within the bounds of our study. We therefore run a constrained optimization in which neither teachers nor headteachers can be reduced by more than 10%.

The results are presented below in Table 15, and shows sizable potential gains from reallocating. Specifically, the optimal allocation would reduce both teachers and head teachers budgets by 10% and spend the resources on Fellows (0.13-0.14) and especially CEVs (1.85-1.89). Although these budget reallocations are modest, the potential gains are sizable, almost approaching the gains of the BT program that come from additional resources. Specifically, the gains relative to status quo are roughly 30% of a standard deviation in the 10th percentile performing school, 26% in the median performing school, and still sizable 23% in the 90th percentile performing school. Importantly, these gains come with only modest budget reallocations and no changes in the existing school-level budget overall.

7 Conclusion

We identify a free lunch for schools in low-income countries. Our model shows how schools can increase student learning without increasing costs by optimally reallocating staff time between tasks. We do this by combining a field experiment that randomly allocates staff of various types with a structural model that enables us to simulate student outcomes under budget-neutral changes to schools' labor structures. The experiment shows that adding new staff to schools alongside a new level of management increases student numeracy by one-third of a standard deviation and leads existing staff to reallocate their time between tasks. We use the results of the experiment to estimate a task-based production function in which managers and workers of different types vary in their comparative advantage across tasks. Importantly, we show that the production function is nonhomothetic in tasks: tasks vary in their relative priority as student outcomes improve. We use this model to simulate budget-neutral counterfactual allocations of staff.

The NGO's staffing allocation is close to optimal given the resources used, but we show how large gains are possible from modest reallocations of *school* budgets toward the new staff roles introduced by the NGO. These budget-neutral changes in the allocation of the school budget lead to gains of about one-quarter of a standard deviation in student learning, with the lowest performing schools gaining the most. The intervention that we study, thus, illustrates how gains in organizational productivity are possible simply from reallocating existing resources, an attractive prospect for schools in low-income countries, which face strained budgets and persistently low levels of student learning. Similar gains may be possible in other public service budgets, especially given that public sector organizations lack a profit motive to achieve organizational efficiency. We leave this for future research.

References

Aghion, P., I. Almås, and C. Meghir (2025): "Human Capital and Development," Working Paper 34602, National Bureau of Economic Research.

Alfaro-Urena, A., I. Manelici, and J. P. Vasquez (2022): "The effects of joining multinational supply chains: New evidence from firm-to-firm linkages," *The Quarterly Journal of Economics*, 137, 1495–1552.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton (2016): "Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India," Tech. rep., National Bureau of Economic Research.

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

Bassi, M., C. Meghir, and A. Reynoso (2020): "Education Quality and Teaching Practices," *The Economic Journal*, 130, 1937–1965.

Bassi, V., J. H. Lee, A. Peter, T. Porzio, R. Sen, and E. Tugume (2025): "Self-employment within the firm," Tech. rep., National Bureau of Economic Research.

Beuermann, D. W., J. Cristia, S. Cueto, O. Malamud, and Y. Cruz-Aguayo (2015): "One Laptop per Child at Home: Short-Term Impacts from a Randomized Experiment in Peru," *American Economic Journal: Applied Economics*, 7, 53–80.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 128, 1–51.

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen (2015): "Does management matter in schools?" *The Economic Journal*, 125, 647–674.

Bold, T., D. Filmer, G. Martin, E. Molina, B. Stacy, C. Rockmore, J. Svensson, and W. Wane (2017): "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa," *Journal of Economic Perspectives*, 31, 185–204.

BT (2025): *Building Tomorrow*, website available at https://www.buildingtomorrow.org, accessed 05 Aug 2025.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 104, 2633–2679.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman (2013): "School Inputs, Household Substitution, and Test Scores," *American Economic Journal: Applied Economics*, 5, 29–57.

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

Dhuey, E. and J. Smith (2018): "How school principals influence student learning," *Empirical Economics*, 54, 851–882.

Duflo, E., P. Dupas, and M. Kremer (2015): "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools," *Journal of Public Economics*, 123, 92–110.

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 102, 1241–78.

Fujimoto, J., D. Lagakos, and M. VanVuren (2023): "Macroeconomic Effects of 'Free' Secondary Schooling in the Developing World," NBER Working Papers 31029, National Bureau of Economic Research, Inc.

Ganimian, A. J., K. Muralidharan, and C. R. Walters (2024): "Augmenting state capacity for child development: Experimental evidence from India," *Journal of Political Economy*, 132, 1565–1602.

GEM (2014): "Uwezo: Monitoring children's competencies in East Africa," *ASSESSMENT GEMS SERIES*.

Hanushek, E. A. and L. Woessmann (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–68.

Hjort, J., H. Malmberg, and T. Schoellman (2025): "The missing middle managers: Labor costs, firm structure, and development," Tech. rep.

Kraft, M. A., W. H. Marinell, and D. Shen-Wei Yee (2016): "School organizational contexts, teacher turnover, and student achievement: Evidence from panel data," *American Educational Research Journal*, 53, 1411–1449.

Lemos, R., K. Muralidharan, and D. Scur (2024): "Personnel management and school productivity: Evidence from india," *The Economic Journal*, 134, 2071–2100.

Muralidharan, K. and V. Sundararaman (2013): "Contract teachers: Experimental evidence from India," Tech. rep., National bureau of economic research.

Nyqvist, M. B. and A. Guariso (2021): "Supporting learning in and out of school: Experimental evidence from India," Tech. rep., Stockholm School of Economics.

Rich, M. (2013): "In Raising Scores, 1 2 3 Is Easier Than A B C," *The New York Times*.

Schoellman, T. (2011): "Education Quality and Development Accounting," *The Review of Economic Studies*, 79, 388–417.

Singh, A., M. Romero, and K. Muralidharan (2024): "COVID-19 Learning loss and recovery: Panel data evidence from India," *Journal of Human Resources*.

TaRL (2025): *Teaching at the Right Level Africa*, website Available at https://teachingattherightlevel.org/, accessed 05 Aug 2025.

Figures

Figure 1: Experimental Design

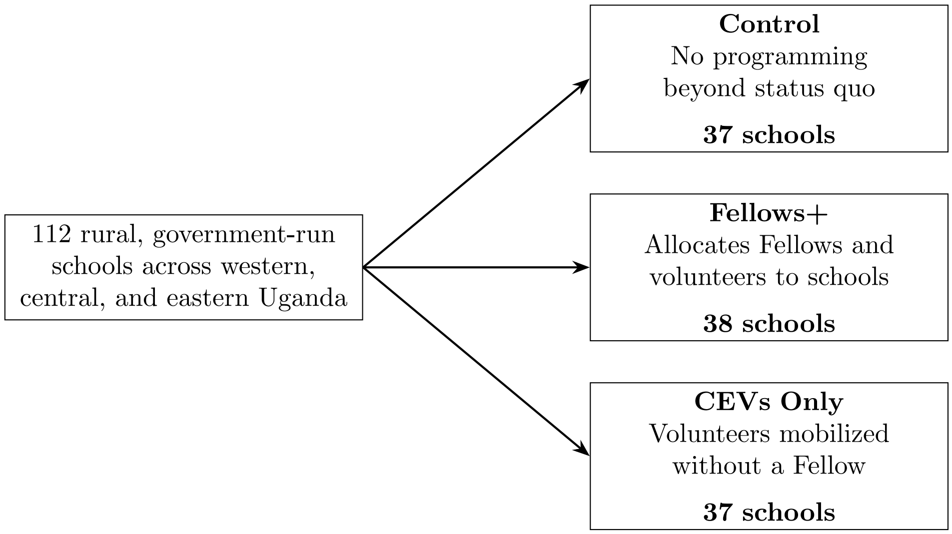
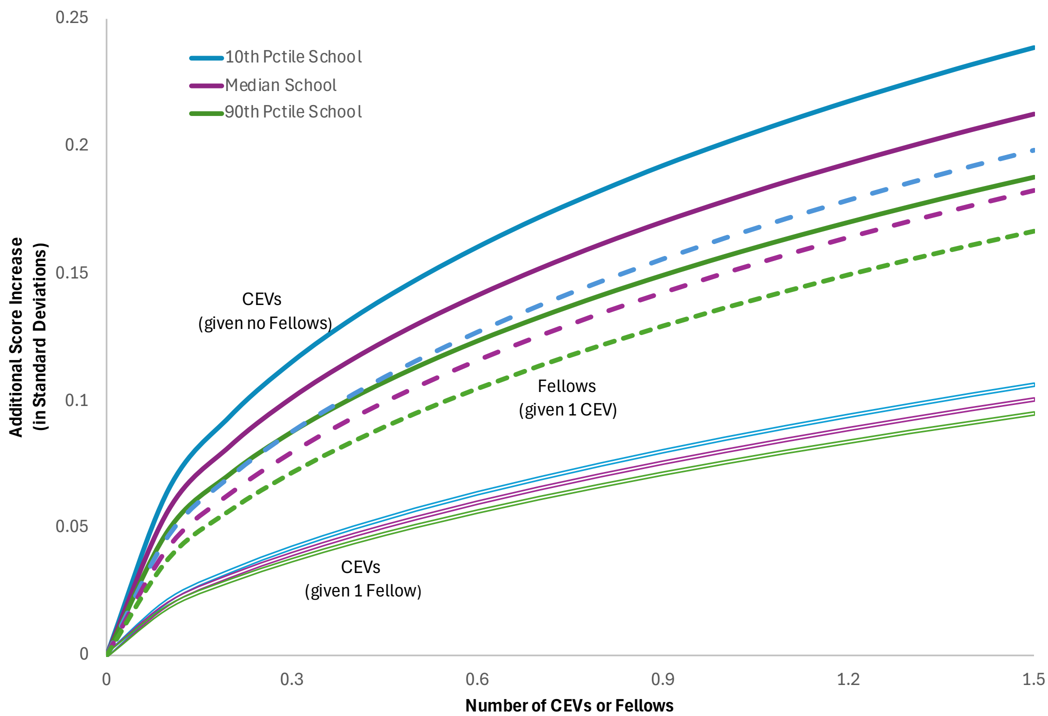


Figure 2: Comparative Statics of Additional CEVs and Fellows on Test Scores



Tables

Table 1: Effects on Assessment Scores, Standard Deviation Units

	(1) Math	(2) Ethnomath	(3) English	(4) Local Language	(5) Numeracy	(6) Literacy
CEVs Only	0.11	0.17*	0.05	0.05	0.18**	0.05
	(0.07)	(0.09)	(0.08)	(0.06)	(0.08)	(0.07)
Fellows+	0.15*	0.36**	0.07	0.04	0.33***	0.06
	(0.07)	(0.13)	(0.07)	(0.08)	(0.10)	(0.07)
Mean Gain in Control	0.369	-0.285	0.587	0.158	-0.006	0.377
Obs.	3,983	3,983	3,983	3,983	3,983	3,983
R^2	0.26	0.23	0.36	0.14	0.19	0.31

Standard errors clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Effects on Threshold Levels

	(1) Add and Subtract	(2) Division	(3) Read Words	(4) Comprehension
CEVs Only	0.06**	0.04	0.04	0.03*
	(0.03)	(0.03)	(0.04)	(0.01)
Fellows+	0.11**	0.05	0.04	-0.00
	(0.04)	(0.04)	(0.03)	(0.01)
Control Mean, Baseline	0.449	0.147	0.359	0.031
Control Mean, Endline	0.592	0.308	0.540	0.053
Obs.	3,983	3,983	3,983	3,983
R^2	0.16	0.19	0.19	0.09

Standard errors clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Effects on Improvement and Regression

	(1) Improve in Numeracy	(2) Regress in Numeracy	(3) Improve in Literacy	(4) Regress in Literacy
CEVs Only	0.05	-0.08***	0.04	-0.01
	(0.03)	(0.03)	(0.04)	(0.02)
Fellows+	0.10***	-0.13***	0.04	-0.01
	(0.03)	(0.03)	(0.04)	(0.02)
Control Mean	0.432	0.440	0.554	0.147
Obs.	3,983	3,983	3,983	3,983
R^2	0.37	0.51	0.44	0.77

Standard errors clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Differential Effects for Female Students

	(1) Numeracy (SD)	(2) Add and Subtract	(3) Improve Numeracy	(4) Regress Numeracy
CEVs Only	0.14	0.04	0.07	-0.10**
	(0.11)	(0.04)	(0.05)	(0.04)
... × Female	0.01	0.05	-0.06	0.02
	(0.09)	(0.05)	(0.05)	(0.04)
Fellows+	0.27*	0.06	0.06	-0.11**
	(0.14)	(0.05)	(0.05)	(0.04)
... × Female	0.18	0.08	0.01	0.00
	(0.14)	(0.07)	(0.06)	(0.06)
Endline Female	0.01	-0.06	0.03	0.00
	(0.09)	(0.05)	(0.03)	(0.03)
Obs.	3,983	3,983	3,983	3,983
R^2	0.09	0.13	0.54	0.81

Standard errors clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Differential Effects for Low-Asset Students

	(1) Numeracy (SD)	(2) Add and Subtract	(3) Improve Numeracy	(4) Regress Numeracy
CEVs Only	0.13	-0.01	0.02	-0.04
	(0.15)	(0.04)	(0.04)	(0.03)
... × Low Asset	0.02	0.12*	0.03	-0.08**
	(0.16)	(0.07)	(0.04)	(0.03)
Fellows+	0.31*	0.04	0.07	-0.07*
	(0.16)	(0.05)	(0.05)	(0.04)
... × Low Asset	0.10	0.09	0.00	-0.05
	(0.17)	(0.07)	(0.06)	(0.05)
Endline Low Asset	0.02	-0.07**	-0.01	0.07***
	(0.12)	(0.03)	(0.03)	(0.02)
Obs.	3,983	3,983	3,983	3,983
R^2	0.09	0.13	0.54	0.81

Standard errors clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Differential Effects for Newly Enrolled Students

	(1) Numeracy (SD)	(2) Add and Subtract	(3) Improve Numeracy	(4) Regress Numeracy
CEVs Only	0.09	0.05	0.02	-0.07*
	(0.11)	(0.04)	(0.04)	(0.03)
... × Newly Enrolled	0.26	0.11	0.09	-0.11
	(0.16)	(0.09)	(0.09)	(0.08)
Fellows+	0.31**	0.08	0.06	-0.09***
	(0.12)	(0.05)	(0.05)	(0.03)
... × Newly Enrolled	0.27**	0.12	0.04	-0.05
	(0.12)	(0.08)	(0.08)	(0.07)
Endline Newly Enrolled	-0.20*	-0.05	-0.04	0.07
	(0.10)	(0.06)	(0.05)	(0.05)
Obs.	3,983	3,983	3,983	3,983
R^2	0.09	0.13	0.54	0.81

Standard errors clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Time Allocation of Teachers Across Tasks

	Control	CEVs Only	Fellows+	p-value of Differences, CEVs Only	Fellows+
Administration	5.8%	7.7%	6.3%	0.10*	0.80
Enrollment	3.6%	2.5%	2.3%	0.45	0.11
Individualized Instruction	5.5%	6.8%	6.5%	0.57	0.54
Classroom Instruction	49.0%	49.2%	47.0%	0.96	0.77
Preparation	30.8%	30.2%	34.3%	0.87	0.52
Truancy	5.3%	3.6%	3.6%	0.09*	0.10

Table 8: Time Allocation of Head Teachers Across Tasks

	Control	CEVs Only	Fellows+	p-value of Differences, CEVs Only	Fellows+
Administration	30.9%	31.2%	31.8%	0.96	0.89
Enrollment	4.0%	5.8%	5.1%	0.23	0.31
Individualized Instruction	9.1%	6.9%	5.6%	0.34	0.09*
Classroom Instruction	25.7%	26.0%	29.9%	0.95	0.44
Preparation	21.1%	23.5%	21.7%	0.55	0.94
Truancy	9.3%	6.6%	6.2%	0.15	0.04**

Table 9: Time Allocation of CEVs and Fellows, by Arm

	Fellows	CEVs, without Fellows	CEVs, with Fellows	CEV Diff., p-value
Administration	6.4%	8.4%	5.0%	0.01***
Enrollment	11.8%	5.2%	2.3%	0.16
Individualized Instruction	59.8%	61.5%	67.7%	0.13
Classroom Instruction	5.0%	0.4%	0.9%	0.04**
Preparation	11.3%	21.0%	20.6%	0.81
Truancy	5.7%	3.4%	3.5%	0.92

Table 10: Sum of Total Hours over All Staff, per Student per Term

	Control	CEVs Only	Fellows+
Administration	0.69	0.97	1.48
Enrollment	0.37	0.30	0.63
Individualized Instruction	0.66	1.45	2.28
Classroom Instruction	5.38	5.73	8.53
Preparation	3.13	3.33	3.95
Truancy	0.42	0.39	0.43
Total Time	10.65	12.17	17.30

Table 11: Model Estimation

Dependent Variable: $\log(l_{ijk}/l_{ij0})$	Non-Homothetic		Homothetic	
	OLS	IV	OLS	IV
Total Task Time $\log(T_{jk}/T_{j0})$	0.60***	-0.42**	0.60***	-0.41**
	(0.08)	(0.20)	(0.08)	(0.20)
Test Scores $\log(\tilde{y}_j)$...				
... × Administration	-0.00	0.01***		
	(0.00)	(0.00)		
... × Enrollment	-0.01***	-0.01***		
	(0.00)	(0.00)		
... × One-on-One	0.00	0.02***		
	(0.00)	(0.00)		
... × Classroom Instruction	-0.00**	-0.00		
	(0.00)	(0.00)		
... × Truancy	-0.00***	0.01***		
	(0.00)	(0.00)		
Task by Worker Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,089	1,089	1,089	1,089
R-squared	0.32	0.16	0.32	0.16
First Stage F-stat		25.70		24.74

Standard errors clustered at the district level in parentheses.
In the IV regressions, the treatment indicators instrument Change in Total Task Time.
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Parameter Values

Variable		Non-Homothetic	Homothetic
Effort Cost Curvature	γ	1.55	3.15
Task Aggregation Curvature	σ	0.77	0.12
Scale Effect on...			
... Preparation	ρ_0	3.14	2.98
... Administration	ρ_1	3.07	2.98
... Enrollment	ρ_2	3.21	2.98
... One-on-One	ρ_3	3.08	2.98
... Classroom	ρ_4	3.08	2.98
... Truancy	ρ_5	3.09	2.98

Table 13: Costs of Each Type of Labor

	Teacher	Head Teacher	CEV	Fellow
Cost in USD per worker	$1,917	$3,058	$170	$11,516
Workers per School (Fellows arm)	7.13	1	4	0.25
Cost per school Relative to Teacher	1.00	0.22	0.05	0.21

Table 14: Budget-Neutral Optimization of Building Tomorrow Program

Teachers	10th percentile	Median	90th percentile
Head Teach	1	1	1
CEVs	3.95	3.99	4.04
Fellows	0.30	0.29	0.28
Gains relative to no program	0.40	0.36	0.33
Gains relative to Fellows+ program	0.04	0.04	0.03

Table 15: Budget-Neutral Re-allocation of School Budget

Teachers	10th percentile	Median	90th percentile
Teacher	0.90	0.90	0.90
Head Teacher	0.90	0.90	0.90
CEV	1.85	1.87	1.89
Fellow	0.14	0.14	0.13
Gains relative to status quo	0.30	0.26	0.23

A Enrollment

Here we examine the impact of both programs on the likelihood that students are unenrolled at endline. Recall that all students in our sample were enrolled in P4 at baseline, and we track students regardless of their enrollment status at endline. We see that the CEVs Only treatment reduces the likelihood of being unenrolled at endline by 2.9 percentage points (p <0.05), a 47% reduction in the likelihood of being enrolled relative to the base rate of 6.2% in the control group. Students in the Fellows+ arm are also less likely to be unenrolled at endline, but the difference is statistically indistinguishable from the control group.

Table A.1: Effect on Likelihood of Being Unenrolled at Endline

	Unenrolled at endline=1
CEVs Only	-.029**
	(.013)
Fellows+	-.017
	(.015)
Control Mean	.062
Observations	1,943

Notes: All specifications include district FE's. Standard errors are clustered at the district level. ***p<0.01, ** p<0.05, * p<0.1

B Matrix of Type-Task Costs

Here we list all the effort cost terms θ_{ik} using the parameters from Table 12 and the formula in equation (5.16) applied to data from the Fellows treatment arm.

Note that these numbers have no independent interpretation, but these can be used to infer comparative advantage by differencing across tasks for the same type, then comparing those differences across types. For example, in the non-homothetic case we can see that the difference in cost for teachers between Classroom and Administration is about 5.4, while the same difference for Head Teachers is about 0.3. From this we can say that Head Teachers have a comparative advantage in Administration relative to teachers (when comparing just Classroom and Administration).

Table B.1: Effort Costs, Non-Homothetic Case

$\log(\theta_{ik})$	Teachers	Head Teachers	CEVs	Fellows
Preparation	-20.3456	-18.1791	-17.6585	-18.8403
Administration	-15.3918	-18.5580	-17.1019	-16.7852
Enrollment	-12.4832	-14.4701	-15.7562	-14.6780
One-on-One	-16.6795	-16.0481	-21.0339	-22.1470
Classroom	-20.7839	-18.8966	-14.7738	-12.9816
Truancy	-15.8163	-14.9947	-15.2066	-13.9022

Table B.2: Effort Costs, Homothetic Case

$\log(\theta_{ik})$	Teachers	Head Teachers	CEVs	Fellows
Preparation	-29.0821	-25.7190	-24.5225	-25.9770
Administration	-20.1488	-26.5019	-23.6009	-22.3252
Enrollment	-14.7320	-18.9363	-21.0317	-18.3686
One-on-One	-22.5521	-21.9859	-30.8223	-32.1518
Classroom	-29.8443	-26.9897	-19.2299	-15.2659
Truancy	-20.9490	-20.0361	-20.1741	-17.0990