THE UNIVERSITY OF CHICAGO

LACKING INCENTIVES, TECHNOLOGIES, OR BOTH? VALUING PRODUCTIVITY THRESHOLDS USING A NOVEL EDUCATIONAL FIELD EXPERIMENT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
HEE KWON SEO

CHICAGO, ILLINOIS
JUNE 2020

Table of Contents

List of Tables

List of Figures

Abstract

Providing students with either "demand-side" incentives (such as performance-based rewards) or "supply-side" materials (such as books) often produces null effects; might there be complementary returns to providing both, and if so, why? In a three-year field experiment with 170 high schools in Tanzania, where mathematics pass rates remain below 20 percent nationally, students were provided with (1) money pegged to math test scores; (2) technologies to ease effort costs of learning; or (3) both of the above. Money or technologies alone make limited impact on test scores, while both together make a large, complementary effect, especially on the scores of students just below the top 20 percent (0.3σ), revealing an inverse-U-shaped relationship between entering grade-level performance and treatment complementarity. I first present a pre-specified benchmark model in which heterogeneous students balance achievement returns against strictly convex effort costs. I then generalize the model to allow minimum interest and knowledge thresholds: students not interested enough or prepared enough to learn have unproductive effort, possibly discouraged by large "entry" costs of learning new material on the curriculum. The thresholds model, structurally estimated based on detailed surveys of study habits matched to outcomes, generates the treatment-effect patterns while the benchmark model does not. Counterfactual simulations suggest that *both* providing the experimental technologies *and* motivating the students by doubling their chance of promotion would induce a modest but meaningful endogenous response of student knowledge, by reducing the share of students who are giving up on learning new material from 79 percent to 52. By explaining treatment complementarities, the proposed analytical framework extends models of classroom learning and welfare implications of policies outlined in the previous literature to reflect a higher degree of realism about developing community contexts. In particular, although the cost of providing technologies through the program is lower than the estimated cost of asking students to attain the equivalent preparation on their own, taking into account the effort costs associated with the "entry" portion of individual learning curves reduces benchmark estimates of the interventions' revealed-preference-based welfare impacts by more than two thirds.

Acknowledgments

Although this dissertation is a single-authored work, several thousands of people helped make progress on this work involving a large-scale field project. These people include my committee members; other faculty mentors and colleagues; field partners; grant and other support partners; family and friends. I would like to acknowledge some of these people by name and others by their affiliation.

I am indebted to my committee Michael Greenstone, Canice Prendergast, Marianne Bertrand, and Michael Dinerstein for their critical advice and support, including Michael Greenstone's recommendation letters to the International Growth Centre (IGC) and the Abdul Latif Jameel Poverty Action Lab (J-PAL). I am grateful to my committee for their generosity with their time and wisdom, and for challenging me to strive for world-class standards of research.

I would like to thank Yuehao Bai, Fiona Burlig, Tamma Carleton, Jonathan Davis, Justin Leung, Leonard Goff, Ali Hortacsu, Amir Jina, James Kiselik, Margaux Luflade, Karthik Nagarajan, Ishan Nath, Chad Syverson, James Traina, Petra Todd, and seminar participants at the University of Chicago, EMCON, World Bank, and University of Pennsylvania; and reviewers at the IGC and J-PAL Post-Primary Education Initiative for helpful discussions.

The field project has involved a collaboration between Zola Electric, Energy Policy Institute at Chicago (EPIC), President's Office of Tanzania – Regional Administration and Local Government (PO-RALG), and Youth Shaping & Sharpening Movement (YSSM). I thank Galt MacDermot for introducing me to Zola, and MinKyung Kim at Centennial Christian Seminary for introducing me to Tanzania's secondary education context. I thank Kervin Bwemelo, Jessica Eastling, Galt MacDermot, Matt Schiller and others at Zola for their solar energy initiatives and support of partnering schools. I thank Sam Ori and Colleen Reda at EPIC for their support during the signing of the memorandum of understanding with the PO-RALG and the processing of the grants. I thank Issa Asesisye, Hadija Maggid, Bernard Makali, Jenimina Mtitu, Jane Nyamsenda, Benjamin Oganga and others at PO-RALG for their administrative leadership, and teachers and students across 170 participating schools for partnership. I thank 40 members at YSSM led by

Elihuruma Maruma for their faithfulness and generosity toward their brothers and sisters in Tanzania.

For financial support, I would like to thank the GivePower Foundation (GP) for the costs of solar facilities, and the IGC, J-PAL and Chicago Booth PhD Program Office for the costs of research; I also gratefully acknowledge tuition and stipend support from the Katherine Dusak Miller PhD Fellowship.

For further technical and administrative support, I would like to thank TETEA Inc. for their open resources; Ingrid Brill at GP; Ani Bhagtiani, Andre Castro, Joshua Chipman, Anne Laski, Rayner Tabetando, Christina Wagner et al. at the IGC; Radhika Bhula, Meagan Neal and Priyanka Varma at J-PAL; Malaina Brown, Cynthia Hillman, Amity James, and Kimberly Mayer at the Chicago Booth PhD Program Office; Social Sciences Computing Services (SSCS) for computational resources and Stephen Mohr for server support; Herb Rhee and staff at Innovative Technology and Education Center for providing office and energy support at unexpected times of need; and Colleen Mullarkey at the University of Chicago Dissertation Office. Errors are mine alone and not due to project partners.

I am grateful to my parents, my sister, my wife Kang Chon Kim, and friends Justin Leung, Solomon David Mbise, and Yosiah Nehemia Mkumbwa. I would not have been able to conduct this work without their undying love and support. I give thanks to God for all the protection over the years, and for all the grace and kindness people have shown me over the course of this research.

1 Introduction

Student achievement of skills is critical to raising living standards across nations, which now near universal primary school enrollment and 75 percent secondary school enrollment (Lee and Lee, 2016). To what extent these classroom hours are productive remains a question of some concern, however; the recent World Development Report documents detailed evidence of classroom learning failures from a wide variety of contexts and, in fact, warns of a global "learning crisis" (World Bank, 2018). Tanzania provides a case in point, where right now less than 20% of high school students pass the national promotional mathematics test.[1]

Past work in education economics has traditionally focused on lack of either "demand-side" incentives (such as motivation and perceived returns) or "supply-side" technologies (such as books, computers and instruction) as potential explanations for substandard achievement—tending to view students as "producers" of knowledge supplying knowledge to their future selves. Yet, a number of experimental investigations in this vein have found null achievement effects of providing either incentives or inputs alone (Fryer, 2011; Kremer, Brannen and Glennerster, 2013).[2] Limited evidence to date exists on whether providing both incentives and inputs can be effective where providing either alone cannot.

This paper reports results from a field experiment conducted with 170 high schools that reached 6,201 students, and a three-year follow-up; students were provided with (1) money pegged to math test scores, (2) technologies to ease the effort cost of learning, or (3) both of the above. Specifically, Group 1 (G1) students were provided with money-reward contracts for marks to be obtained on a year-end curriculum-based test ("Incentives"). Group 2 (G2) students were provided inputs combining free solar-energy access, bilingual textbooks, and videos ("Technologies") that were designed to better show students not just what but *how* to study—based on Glewwe, Kremer and Moulin's (2009) suggestion that a mismatch between

[1] About 40% of age-group youths sit as candidates, which means that among Tanzania's 10 million men and women aged 18 to 27, only 6% have passed high school mathematics; passing mathematics is a prerequisite to training in STEM-related occupations, and hence may be high stakes for some students (NECTA, 2019).

[2] Even in non-experimental works, reported effects are often null, though more debate exists (Hanushek, 2020).

one's mother tongue and the language of instruction may hinder learning, and Fryer's (2011) suggestion that students not knowing how to study may also hinder learning. Group 3 (G3) students were provided both of the above ("Both"), testing the cross effect of the G1 and G2 intervention components directly delivered to students. Each treatment was delivered in the beginning of each year, and the same treatment groups were followed for a period of three years, between 2016 and 2018.

Results show that money or technologies alone make limited impact on test scores (0.05 and 0.09σ, respectively), while both together make a large impact (0.33σ for the both treatment effect and 0.19σ for the interaction effect), especially on the scores of students just below the top 20 percent (0.35σ for the both treatment effect and 0.33σ for the interaction effect).[3] Hence, the interaction effects display an inverse-U-shaped relationship between entering grade-level performance and treatment complementarity. Detailed surveys of study habits matched to outcomes point to increased hours of study and higher levels of attention during study, particularly with the provided books, as the mechanisms for improvement. The results raise the question about how to understand the economics reasons behind them: why does providing both of these interventions lead students to substantially improve effective effort where providing either alone does not, and why does the interaction work more strongly for some students than for others?

In order to rationalize these results, based on the surveys of study habits matched to outcomes, I estimate a model of students who recognize the benefits and costs of learning that generates key patterns of treatment effects. I begin from a benchmark model of classroom learning with standard agency assumptions: students recognize the benefits and costs of practicing their mathematics skills to improve their test scores (Todd and Wolpin, 2018). I then hypothesize a simple additional feature: students face certain "entry" costs of acquiring new mathematics skills—even before they are able to practice new skills—that they incur only if they see suffi-

[3]The effects reported are from the second year of the program, as shown in column 6 of table 3; column 9 of the same table shows that year 3 results are similar. The "interaction" effect refers to the G3 (both) treatment effect minus G1 (incentives) and G2 (technology) effects.

ciently large benefits. I test for the presence of these "productivity thresholds": levels of endowments that must be crossed before students can begin improving their test scores. In particular, I estimate minimum-interest threshold and minimum entering-grade-level-knowledge (or "preparedness") threshold.[4] I show that this feature helps explain key treatment-effect patterns, in particular the hump-shaped relationship between the interaction effect and the student's pretest score that the standard model without thresholds costs cannot explain. I show that these thresholds are consistent with a model of students facing large "entry" costs of learning new material on the curriculum. Since these models give different policy implications, selecting the model that offers greater fidelity to empirical realities has direct implications for policy effectiveness, which I assess in counterfactual simulations.

To elaborate, the models enable me to value test scores and promotion; compute welfare implications of the interventions; and simulate counterfactual outcomes from lowering the promotional cutoff—a low-cost policy option that would offer more students a realistic chance at promotion and, therefore, a higher expected return from marginal effort. The results from the productivity-thresholds model in particular suggest that, conditional on providing the experimental technologies, lowering the cutoff will generate a modest but meaningful endogenous response of student knowledge, foremost by reducing the share of students who are "giving up" on learning new materials—i.e., students not putting in the threshold costs of acquiring new skills—from 79% to 52%. However, taking into account the implied effort costs associated with climbing the "entry" portion of individual learning curves reduces benchmark estimates of the interventions' revealed-preference-based welfare impacts by more than two thirds, and critically lowers the policy effectiveness of these types of interventions.

In doing so, I demonstrate how to use field-experimental treatment variations to estimate the parameters of a student-learning model based on standard agency assumptions, and ad-

[4]These thresholds may be viewed as a parsimonious way of capturing complex micro-foundations that may be at play in the classroom-learning setting, such as learning-by-doing, habit formation, or heterogeneous learning curves (Loerch, 2001). Characterizing threshold/fixed costs using a structural model has also been of interest broadly across contexts of development, such as Chilean manufacturers (Levinsohn and Petrin, 2003), Indian farms (Foster and Rosenzweig, 2017), and Kenyan agricultural intermediaries (Bergquist and Dinerstein, 2019).

ditionally identify productivity thresholds that stand in the way of learning. While hundreds of field experiments have examined classroom-level test scores, and an extensive body of work has studied skill production functions, few models have focused on student self-agency: that is, students explicitly choosing their own levels of effort and performance in the classroom. My work provides an intuitive, clarifying measure of classroom efficiency from this perspective: I measure the percentage of students who meaningfully accumulate knowledge. Conversely, this share is one minus the proportion of students who seem from the data to be turning off their minds during the year lacking the preparation and motivation to follow the material.

Indeed, to my knowledge, this work is the first in the literature to use field-experimental treatment variations to estimate a model that explicitly characterizes the benefit-cost consideration of *students* (as opposed to parents), enabled by multiple recent advances from three strands of the economic literature. First, this work builds on insights from recent field experiments in education involving incentives (Fryer, 2011; Hirshleifer, 2017), complementarities (Behrman et al., 2015; Mbiti et al., 2019), and teaching at the right level (Duflo, Dupas and Kremer, 2011; Banerjee et al., 2016). Second, this work builds on the technology of skill formation literature that tends to approach skill production functions with a more agnostic (less explicitly agency-based) stance on learning (Cunha and Heckman, 2008; Cunha, Heckman and Schennach, 2010; Agostinelli and Wiswall, 2016). Third, this work builds on a recent exercise of structural estimation of classroom learning that endogenizes a key choice variable in the process of skill formation: student's own effort (Todd and Wolpin, 2018).

Adding to these works, this paper provides the first experimental test of complementarity between interventions provided strictly to students as opposed to teachers. It delivers valuations of test scores that cannot be inferred from treatment effects alone, since these effects confound the willingness-to-pay (WTP) for knowledge with the cost of effort. This paper also points to threshold costs of investment as a source of input complementarities that get reported to be significant in estimates of educational production functions. Finally, this paper provides the first case study using field-experimental treatment variations to assess structural parameters

of classroom learning with student self-agency, which may be of particular relevance to communities where the number of first generations of secondary-school attendees—with limited parental experience in education—is beginning to rise. By providing a measure of classroom efficiency from this perspective, I argue that the work can be used to inform the targeting of educational investments and curriculum design.[5] For the context of Tanzania in particular, I ask: (1) whether the high bar of the test is constraining the efficiency of mandatory hours students are being asked to spend in the mathematics classroom; (2) why providing additional inputs alone or lowering the bar alone without providing additional inputs is expected to be fruitless; (3) how, conditional on providing more inputs, the bar may be better targeted; and (4) why large responses in test scores may not imply large changes in welfare.

This paper proceeds as follows. Section 2 reports the details of the field experiment and reduced-form results. Section 3 outlines the conceptual framework. Section 4 describes the structural-estimation framework and its results. Section 6 concludes.

2 Experiment and Reduced-Form Findings

2.1 Setting

The "Sharpening Mathematics Review (SMR)" project began in the beginning of 2016, partnering with 9th-grade students in 170 rural Tanzanian high schools, President's Office – Regional Administration and Local Government (PO-RALG), and seven other organizations.[6] The SMR project involved a collection of interventions developed to support mathematics education of secondary school students. The sample of schools selected were the population of all schools without electricity in 23 northern Tanzanian districts that were enlisted in September,

[5]Students in Tanzania's junior secondary schools—approximately 1.5 million in number—are mandated to spend at minimum 100 hours of classroom-learning on mathematics each year, more so than on any other subject.

[6]These included Zola Electric; GivePower foundation; Energy Policy Institute at Chicago; Youth Shaping & Sharpening Movement; International Growth Centre; Abdul Latif Jameel Poverty Action Lab; and Chicago Booth PhD Program Office.

2015.[7] The districts selected were the intersection of where Zola Electric, a national energy company and research partner, was servicing at the time, and where the government deemed performance to be relatively low in terms of graduation rates.

The SMR sample represent a "middle class" of students across the nation. Figure 1a shows O Level aggregate and selected subject-level pass rates nationwide between 2012 and 2015. Out of 1.4 million students who sat for these examinations across these four years, 55% of students managed to obtain O Level certification, which requires getting at least two D's or above on subject-level examinations out of seven best subjects taken.[8] It can also be seen, re-scaled on the right-hand-size y-axis, that the same number re-scaled corresponded to 23% out of 3.4 million youths in the official age group population. That is, less than a quarter of youths in the junior-secondary age group were able to obtain certification over these years. Figure 1b shows the corresponding figures in sample schools, where the pass rates are generally lower, reflecting the selection criterion of relatively lower performance. The average pass rate in the selected sample schools corresponded to the bottom 25th percentile of nationwide school pass-rate distribution. Note that, as can be seen in fig. 1a, if a student is enrolled in secondary-school, the student is already in the top 40% of the nation in terms of educational attainment. Hence, this group of students could be seen as representing a "middle class" of students in terms of educational attainment among the age-group population.

Secondary mathematics has been of particular interest to the government because the subject-level performance has remained poor. It can seen that 40 to 60% of students passed Swahili, English and Biology, whereas the pass rate for mathematics stood below 15% nationwide. Among the sample schools, the pass rate for mathematics has stood below 6%.[9]

[7] The project initially identified 173 schools, but two schools closed as the program was beginning in 2016 and another school was disqualified from government examinations that year because of irregularities. Data were not collected from these schools.

[8] O Level involves five required subjects and two or more optional subjects; the five required are Swahili, English, Biology, Civics and Mathematics. A grade of D means that the student has obtained at least 29.5 marks out of 100 marks total on the subject level examination. Civics rates are omitted in the figures to conserve on space, but the rates are similar to those for Swahili and English.

[9] Re-taking the test requires repeating at least two years of schooling or attending a private school, both of which are cost-prohibitive for most adolescents in the nation. Re-takers also lose post-O-Level public-tuition support for which regular candidates qualify.

2.2 Design

The SMR project involved 170 classrooms from 170 schools, one randomly selected ninth-grade classroom from each school. Between 2015 and 2016, these classrooms were randomly divided into four groups: Incentives (G1), Technology (G2), Both of the above (G3) and Control (C). Incentives provided cash to students for scores on year-end mathematics tests. Technology provided solar-energy access, bilingual textbooks and videos, with emphases not just on what to study but how to practice.[10] The treatments were delivered every year for three years, with some variation between years, as detailed below. The details, including the implementation timeline, can also be found diagrammatically summarized in figs. B1 and B2.[11]

- **Year 1**: Schools were randomized into 3 treatment groups and 1 control group, with some additional variation in incentive-contract amounts within each classroom.

  G1 ("Incentives Only"): Students were given a fixed piece-rate incentive contract for each mark to be scored on an end-of-the-year curriculum-based mathematics test (42 schools).

  G1.1 1/4 of students were promised $0.125 per mark.

  G1.2 1/4 were promised $0.25 per mark.

  G1.3 1/4 were promised $0.50 per mark.

  G1.4 1/4 were promised $0.75 per mark.

  G2 ("Technology Only"): Students received 9th-grade mathematics textbooks with Swahili chapter summaries. Schools received solar panels (covering approximately two large classrooms and one office), two TVs (one 16-inch and one 19-inch), and a set of 15-hour mathematics videos covering the full 9th-grade curriculum (44 schools).

  G3 ("Both") Students received both Incentives and Technology (44 schools).

[10] Angrist and Lavy (2002) and Cristia et al. (2017), for example, find null effects of providing computers to schools in Israel and children in Paraguay, respectively.

[11] Currently under preparation is a draft photo essay depicting some of these activities: link.

C ("Control") Students received neither Incentives nor Technology (40 schools).

- **Year 2**: Same treatments continued in year 2, except with the incentive contracts unified.

  G1 ("Incentives Only") Students were promised $0.50 per mark.

  G2 ("Technology Only") Students were given 10th-grade textbooks and videos.

  G3 ("Both") Students received both Incentives and Technology.

  C ("Control") Students received neither Incentives nor Technology.

- **Year 3**: Same treatments continued in year 3, except schools that had been without solar also received solar in the beginning of the year. Hence, the Technology variation only involved textbooks and videos in year 3.

  G1 ("Incentives Only") Students were promised $0.50 per mark.

  G2 ("Technology Only") Students were given 11th-grade textbooks and videos.

  G3 ("Both") Students received both Incentives and Technology.

  C ("Control") Students received neither Incentives nor Technology.

The equalization of piece rates in the beginning of year 2 was for the concern that invidious effects of comparison might hurt the intrinsic motivation of some students.[12] The equalization of the solar variation in the beginning of year 3 was to honor an initial agreement with the government that non-receiving schools would receive the same solar facilities within two years.[13]

Figure 3 shows photographs from the field capturing some typical deliveries of the interventions. Figure B3 provides linked access to online copies of the program's textbooks and videos. Table A2 provides initial and realized power calculations (c.f. table A1 for reference estimates from past works).

[12] I thank an anonymous JPAL referee for recommending this change. There were a number of anecdotal reports in year 1 that some students were disappointed with this variation even after being informed that the variation was random. See section 2.5 for further discussion.

[13] The treatment effects between years 2 and 3 remained similar, suggesting that solar was not the binding technology.

In February, 2016, YSSM field team conducted the Form 1 (grade-8) SMR survey and examinations, and immediately after the survey and examinations, the field team signed the incentive agreements with the students and distributed the textbooks and videos.[14] Only students who were present during this visit on an arbitrary weekday were enrolled in the program. While participation was voluntary, all students agreed to participate. I take the data from February, 2016, as predetermined ("Year 0") student characteristics.[15]

In October, 2016, YSSM field team conducted F2 (grade-9) survey and examinations ("Year 1 Test"), whose data constitute end-of-first-year observations.

In November, 2016, NECTA conducted the promotional FTNA examination, whose results provide an auxiliary set of observations for the sample students.

In February 2017, the project reinforced a similar design, but with the piece rate equalized across students at $0.50 per mark.

In October, 2017, YSSM field team conducted F3 (grade-10) survey and examinations ("Year 2 Test"), whose data constitute end-of-second-year observations.

In February, 2018, the half of the chools that had not received the solar facilities received the same solar facilities as the other half. The rest of the treatment variation continued; therefore, in year 3, the "Technology" variation consisted only of textbooks and videos.

In October, 2018, YSSM field team conducted F4 (grade-11) survey and examinations ("Year 3 Test"), whose results constitute third year outcomes.

In November, 2018, the sample students took their O Level examinations.

To recap the data, this study relies primarily on survey and test results from February, 2016 (year 0); from October, 2016 (year 1); from October, 2017 (year 2); and from October, 2018 (year 3).

[14] As the PO-RALG was a direct partner, all tests and surveys were known as President's Office Mathematics Evaluation (POME) tests and POME surveys; hence, though non-binding, the instruments of evaluation were as authentic as binding administrative examinations.

[15] Because of field partners' expense schedules, Technology groups (G2 and G3) received pilot installations of solar facilities (for one classroom) toward the end of 2015, before the textbooks, videos, and incentive contracts were announced. See Seo (2016) and Seo (2017) for additional details. The pilot solar exposure in October and November of 2015 did not affect the students' mathematics learning (December is a vacation month for these students). In table 1, year 0 test scores are balanced on the treatment indicators.

2.3 Reduced-Form Estimating Equation

I report difference-in-means estimates, based on the following equation:

$$Q_{ijt} = \beta_t^0 + \sum_{g \in \{1,2,3\}} \beta_t^g \times G_i^g + X_{it}\zeta + \epsilon_{ijt}, \tag{1}$$

where i indexes students; j, schools. The index t denotes evaluation year (the year end).[16] The parameter β^0 is a constant term. Q_{ij}, the explanatory variable of interest, may only exist at the school level (Q_j), in which case the analogous school-level regression is examined. β^G represents the treatment effect of group G. G_i indicates the group to which i belongs. ζ is the vector of coefficients on i's covariates, X_i, which can include age, commute distance, pretest Z-score or a flexible polynomial in attendance propensity score.[17] Because of non-response, controls were missing at random for approximately 10% of observations.[18] Results are robust to multiple (stochastic) imputation, as discussed in detail in section 5. Attrition controls are also discussed in detail in section 5. I cluster standard errors at the school level, the level of the independent unit of the randomized draw (Bertrand, Duflo and Mullainathan, 2004). I report both unadjusted significance levels, and Benjamini, Krieger and Yekutieli (2006) sharpened two-stage q-values (adjusted for three hypotheses) as described in Anderson (2008).

2.4 Reduced-Form Results

Table 1 reports means of student characteristics and tests of their balance. It can be seen in panel A of table 1 that students were on average 15 years old and 56% female at the onset of the evaluation. Less than a quarter of primary guardians completed secondary school. Over two thirds of the parents engaged in farming or fishing, while less than a tenth of the parents engaged in technical or managerial occupations.[19] In contrast, only 1% of students desired

[16] Results are robust to whether I use difference-in-difference specifications.
[17] X_i also includes randomization-block (five-region) indicators.
[18] That is, the missingness of each control was balanced on treatment indicators.
[19] Occupations were categorized by ISCO classifications during data entry based on free responses.

to engage in farming or fishing; 93% of students desired technical or managerial occupations. Note that, as alluded to in section 2.1, approximately half of these students were expected to either drop out or fail the junior secondary pass examinations within three years in spite of these hopes.

In panel B of table 1, it can be seen that 75% of students reported their "intended area of focus" to be Science (as opposed to Arts or Commerce, the two other tracks that students can elect to follow starting in grade 10). Despite such a stated preference for the Science track, only 5% passed the mathematics evaluation in February, 2016 (panel C). Due to random chance, age and commute distance were not balanced during randomization. Hence, I check the robustness of subsequent results to including these variables as controls. Finally, the last row shows that the joint orthogonality of treatment indicators cannot be rejected, demonstrating acceptable overall balance.

As can also be seen in the balance tables, not every student responded to all questions, and for approximately 10% of students, three key demographics variables are missing: age (4.13%), commute distance (7.66%) and pretest score (0.06%). The missingness of these variables was balanced on treatment indicators (regression results omitted, here, though available upon request). I use multiple imputation to ensure robustness of results to missing data.[20] All results are robust and implications do not change if I repeat the same analyses just dropping all observations with missing controls.[21] As discussed next, although controls were missing at random, attendance on subsequent examinations was not, calling for another method to deal with selection not at random.

Table 2 tests whether absences from the mock tests were balanced. Absences were substan-

[20] I use Stata's "multiple imputation by chained equations algorithm" to generate 10 data sets, imputing five variables (age, pretest score, commute distance, math-study hours, and other-study hours), while involving seven auxiliary variables (female, STEM-occupation intended, class size and four region indicators). I then estimate regression models on each filled-in data set, using Rubin's (1987) formulas to combine the parameter estimates and standard errors into a single set of results. According to Enders (2010), multiple imputation is currently regarded as a "state of the art" missing data technique when data are missing at random; not only does it improve the accuracy and power of analyses relative to other techniques, but it also gives full consideration to every sample observation.

[21] Compare tables 3 to 5 and tables A3 to A5. In fact, a number of key treatment effects rise in significance levels when observations with missing controls are simply dropped.

tial and selective. In year 1, as shown column (1), 80% of students showed up on the date of the follow-up evaluation. All three treatment groups (G1-G3) saw higher attendance than the control group. It is noteworthy that even though students in the Technology Only group (G2) were not promised any money rewards, more students showed up on the test day also from G2, suggesting that the program study aids were able to encourage students to at least show up more to the evaluation in year 1. The statistical significance of this effect disappears from G2 in subsequent years, however. In year 2, significantly more students showed up only from G1 and G3; in year 3, significantly more students showed up only from G3.[22] Aggregate attendance fell to 67% and 66% in years 2 and 3. The especially sharp fall between years 1 and 2 was due to students no longer being enrolled in the same grade; reasons for no longer being enrolled—omitted from this table because of space constraints—include failing the 9th-grade promotional examination (12%), quitting school (6%), and transferring to a different school (4%).[23]

Columns (4) and (6) show that not only attendance on test day but also formal enrollment in school was higher among incentivized groups, in line with evidence from past literature that incentives raise enrollment (Barrera-Osorio et al., 2011). These higher enrollment statistics did not translate into higher rates of students graduating, however, showing that students incentivized to stay in school at the margin fail to complete schooling when a stringent graduation test stands in the way.

The substantial and selective attrition motivates the need to control for attrition. I employ the non-parametric control-function approach of Heckman (1990), using commute distance to instrument for attendance. I use the probit analogues of columns (1), (3) and (5) to form selection propensity scores for each student and year, and include a third-degree polynomial of these propensity scores alongside age and pretest scores as controls for selection.[24] Going forward, I report results from this "selection-corrected" specification, in addition to results from (1) specification without any controls, and (2) specification with the standard controls (pretest score—

[22] All regressions control for age, pretest score and commute distance; results are robust when the controls are dropped.

[23] Unlike the O Levels, the 9th-grade promotional examination can be retaken by repeating the grade.

[24] Results are robust to using polynomials of degrees one to ten.

which is standard in education economics—as well as age and commute distance—variables of imbalance). Even though commute distance is my theoretically-preferred instrument for selection, the variable was not balanced because of random chance. Yet, the results are robust not only to this check but also to various other imputation-based checks.[25]

Table 3 reports achievement impacts, also plotted in fig. 4a. In year 1, the estimated effects were noisy though suggestive: the highest coefficient ranged between 0.123σ and 0.183σ for the Both group.[26] In years 2 and 3, the treatment effects were large and highly significant only for the Both group: the coefficients ranged between 0.280σ and 0.415σ for the Both group, while being much weaker and insignificant for the other groups. Since weaker students were more likely to be absent, selection correction was expected to adjust the estimate upward if only the students who were present from the treatment groups were considered; however, selection correction also accounts for potentially weaker effects from absent students across the board (Heckman, 1990). In aggregate, controls and corrections end up attenuating the coefficients, but in negligible manners as seen. Panel C reports linear combination of treatment effects; in particular, $\beta^3 - \beta^2 - \beta^1$ tests the complementarity between interventions G1 and G2. It is shown that the study was marginally underpowered to detect the complementarity effect in this setting, although the estimated magnitudes are meaningful in years 2 and 3, ranging between 0.131σ and 0.186σ. The complementarity effects are significant in alternative specifications that drop observations with missing controls; see table A3. These estimates are comparable to those from past works that have reported various different combinations of "Incentives Only," "Technology/Inputs Only" or "Both" treatments targeted at students across various settings. As seen in table A1, results from past works, when collated together, suggest that the complementarity effect may be large. This work provides, to my knowledge, the first cleanly identified

[25] Other checks include imputing zeros, imputing means, predicting outcomes based on pretest score percentiles, and Lee (2009) bounds. These checks are available upon request.

[26] This pattern of first-year results being weaker than subsequent-year results was also seen in Mbiti et al. (2019), who studied responses of teachers to performance-based incentives in the Tanzanian primary school context and suggested that this was because the teachers did not believe in the first year that the incentive promises were real. See section 2.5 for additional discussion.

evidence on the magnitude of potential complementarity between student-level inputs.[27]

Table 4 examines effects on reported hours per week of mathematics study. The impacts are aligned with those on achievement in that the impacts show up only for the Both group. Whereas the treatment effects were weak and insignificant for G1 and G3, the effects were strong and significant for G3, ranging from 1.0 additional hour per week in year 1 to 2.3 additional hours per week in year 3.[28] Although the complementarity effects are somewhat noisy, the effects show up in economically meaningful manners across specifications in years 2 and 3: on the order of 1.3 hours per week, against the control mean of approximately six hours per week (c.f. table A4).

Table 5 reports performance effects in year 2 and year 3, disaggregated by pretest performance quintiles, following Glewwe, Kremer and Moulin (2009).[29] The year-2 heterogeneous effects are also plotted in fig. 4b. The interaction effects display a hump-shaped pattern across the pretest quintiles, as reported in the bottom row of table 5 and also plotted in fig. 4c. In both years 2 and 3, the magnitudes rise from an essential zero for the bottom two quintiles to reach approximately a third of a standard deviation for the fourth quintile, falling back down somewhat and becoming noisy for the top quintile. This pattern suggests that a significant mass of students toward the upper-middle of the pretest performance distribution under-perform particularly because they lack not only requisite technologies but also motivation to learn. The effect attenuates and becomes noisy for the top quintile, suggesting heterogeneous responses in this quintile (c.f. table A5). It happens to be the case that the promotional cutoff, located at approximately 90th percentile of the performance distribution, stands to be the most relevant to this group; anticipating, I test the empirical content of this relevance in sections 4 and 5.

Two other patterns that may stand out to some observers are how (1) the bottom quintile of

[27] A clean identification of a complementary effect requires a full-factorial (fully interacted) experimental design; as reviewed by (Mbiti et al., 2019), such a design been rarely seen in the education economics literature, because of its difficulty of implementation.

[28] The effects on G3 hours are significant even in year 1; however, G3 performance effects are too noisy to reject the null in year 1.

[29] Due to space constraints, year 1 results are omitted; year 1 results show trends that are similar but weaker and less precise than those from the other years.

students seem to respond to money alone, and (2) the top quintile of students seem to respond to technology alone. Anticipating the discussion in section 5, the results of structural estimation would suggest that (1) is largely due to sampling variation and (2) is largely due to the top quintile of students beginning the grade better prepared and facing stronger motivation.

Table 6 reports on proxies for usage of material/other inputs: (1) hours of study in school after 6pm (proxying for usage of lights), (2) printed-materials usage (proxying for usage of books), (3) ICT and multimedia usage (proxying for usage of videos), and (4) total teacher hours. Effects are generally null for variables (1) and (4), suggesting that students may be able to find other means of sourcing light in school (kerosene, lamps, etc.) and that teachers do not change teaching effort in response to the treatments. Effects are large and significant for variables (2) and (3), but not aligned with the performance effects in that they show up strongly for G2 also and not just for G3. I take these pieces of evidence as justifications for some of my modeling choices: effective student effort can be summarized by a uni-dimensional measure (e.g. hours of mathematics study); technologies can be summarized by a technology parameter; and teacher effort is orthogonal to the treatment interventions studied.

Table A10 reports effects on O Level mathematics certification examination (CSEE) grades and breakdown of the difference in results from that of year 3's mock test. In contrast to the mock test results, whose values can range continuously from 0 to 100, CSEE grades are reported only in grade brackets: 0 to 29.5 marks translate into F; 30 to 44.5 marks into D; 45 to 64.5 marks into C; 65 to 74.5 marks into B; and 75 to 100 marks into A. For each student, I take the midpoint of their corresponding grade bracket range and convert these midpoints into Z-scores. Column (1) shows the result of regressing these Z-scores on the treatment indicators. No significant effect is seen.

The null effects stand in contrast to results seen on the mock test, shown in column (2), bracketed and reconverted to Z-scores in the analogous way. On the mock test, the Both group shows 0.166σ higher test scores, but on the O Level test, this difference reduces to null, suggesting that 40 percent of the difference between effects on the mock test and effects on the O Level

test is attributable to lower resolution of the O Level test's grade brackets (given that at higher resolution the difference was 0.28σ). Columns (3) and (4) show this difference disaggregated into improvements among students whose scored higher on the O Level than on the mock test, versus deteriorations among students who scored lower on the O Level than on the mock test. Column (3) shows that approximately 20 percent of the difference between columns (1) and (2) is attributable to control students catching up, while 40 percent is attributable to erasing of knowledge gains from the both treatment group within the month-long interval between the mock test and the O Level test.

2.5 Validity of Reduced-Form Results

In this subsection, I discuss some threats to validity posed by (1) possibilities of cheating, (2) unusual differences between effects in year 1 and subsequent years, and (3) the difference between the mock test results and the O Level test outcome.

As for (1), cheating was not a concern in this study, as the mathematics examinations were difficult to cheat on: all questions were in free response format, and marks were given only to those students who show valid steps. In year 3, following Mbiti et al. (2019), I randomly provided five different versions of the examination to students. I did not inform the students that the test versions were varied. As seen in table 3, treatment effects remain essentially equivalent across years 2 and 3. I asked markers to indicate any student who attempted to copy answers from a version different from the student's own; less than 0.1% of students attempted to copy.

As for (2), I consider a few reasons why year 1 did not see any significant achievement gain while subsequent years did.

The first is that students may not have found the incentive contracts to be credible in year 1. This pattern of first-year results being weaker than subsequent-year results was also seen in Mbiti et al. (2019), who provided cash incentives to teachers in Tanzanian primary schools; they hypothesized that this was because teachers in the first year did not believe that cash would actually be provided in the first year. I hypothesize a similar reason in this setting. I find mixed

evidence on this hypothesis. On the one hand, since students in incentives-support groups attended the tests with significantly higher probability, it could not have been the case that they thought the probability of payout was zero. Also, lacking belief in the contract is inconsistent with spending significantly higher number of hours on mathematics, as reported among the Both group. On the other hand, I find suggestive evidence across columns (1) and (6) of table A6 that, in year 1, the highest piece-rate groups did not attend the test at a significantly higher rate, while in year 2 they did, possibly suggesting lagged effects of "seeing is believing."

The second is that aversion to inequality in piece rates in year 1 may have discouraged some or all students in year 1; invidious effects of comparison may have hurt the intrinsic motivation of some students and demotivated them from learning. In columns (5) of table A6, I see that a nontrivial portions of students report that the differential piece rates were "demotivating" (as opposed to "motivating" or "did not matter"), and in column (9) I see that almost a fifth of students at least recognize that the piece rate promises were "unfair." Perhaps relatedly, I see in columns (3) and (8) that test scores were generally lowest in groups promised the lowest piece rate, though estimates are noisy.

The third is that the test in year 2 was too difficult, or relatedly, that the textbook in year 1 might not have been as well aligned with the test in year 1 than in subsequent years. The circumstantial evidence for this hypothesis is simply that the Both group studied for significantly more hours in all years, but the treatment effects are large and significant only post year 1.

All of these reasons (as well as potentially others not discussed) may have contributed to the treatment effect being essentially null in year 1. For the structural estimation, I focus on estimating the model in year 2, taking year 0 and year 1 responses as predetermined characteristics.

As for (3): again, I do not see a large or meaningful effect on the actual O Level promotional mathematics test (0.02σ in test score and 0.1 p.p. in pass rate), even though the latter test covered equivalent curriculum subtopics (see table A7) and took place only a month after the

incentivized mock test administered in October of the third year.[30] As discussed, the government only provides access to five letter-grade brackets of the O Level test; less than 20 percent of students receive anything other than "F" on the O Level test, scarcely leaving any room for improvements to be observed for the vast majority of students. The findings highlight how difficult it can be to measure performance when the resolution of the measuring instrument is low; the findings also highlight the difficulty of the test faced by students in this setting, and the limits of what policies can do to help students in such a setting.[31] Nevertheless, these concerns do not threaten the main finding of performance gains caused by the interventions seen in the second and third year of the program.

3 Conceptual Framework

I present a model of students who choose optimal effort, given (1) utility benefits to knowledge and (2) costs of effort, including (i) costs governed by the parameters of the knowledge production function that vary with effort and (ii) threshold costs that can be motivated from learning-curve considerations.

3.1 The Knowledge Production Function

Let i denote a student taught by teacher j in year t. Student i chooses her level of effort, E_{it}, taking as given parameters that govern the productivity of her effort: $K_{i,t-1}$, her initial level of knowledge; A_{jt}, the ability of her teacher; R_{jt}, the effort of her teacher; δ_i, the "regression" rate of her level of performance from the previous grade; and τ, the efficiency (or "factor productiv-

[30] Links to online copies of these tests are provided in fig. B3. NECTA has been aware of the SMR project, and SMR mock questions were submitted to NECTA prior to test administration in all years.

[31] While previous works have reported that teacher impacts on test scores can "fade out" rapidly in subsequent grades, this paper uniquely shows that achievement gains accumulated over three years can fade out as rapidly as in one month.(Banerjee et al., 2007; Rothstein, 2010; Carrell and West, 2010; Jacob, Lefgren and Sims, 2010; Behrman et al., 2015; Mbiti, Romero and Schipper, 2019). Chetty, Friedman and Rockoff (2014) find, in their particular data, "fade-out and re-emergence" effects, whereby teachers' impacts on earnings are similar to predictions based on the cross-sectional correlation between earnings and contemporaneous test score gains, echoing findings of early childhood interventions (Deming, 2009; Heckman, Pinto and Savelyev, 2013).

ity") of commonly supplied educational technology (e.g. textbook).[32] I allow δ_i to vary linearly with $K_{i,t-1}$ to reflect the notion that, for example, when the curriculum content is changing fast year to year, better performing students in the previous year have more performance level to lose by not studying this year, especially where test scores can commonly be zero.

The amount of knowledge student i comes to possess at the end of year t is given by,

$$\begin{aligned} K_{it} &= (1-\delta_i)K_{i,t-1} + \left(\tau A_{jt}^{\gamma_0} R_{jt}^{\gamma_1}\right) K_{i,t-1}^{\alpha_0} E_{it}^{\alpha_1} \\ &= (1-\delta_i)K_{i,t-1} \times \left[1 + \left(\frac{1}{1-\delta_i}\tau A_{jt}^{\gamma_0} R_{jt}^{\gamma_1}\right) K_{i,t-1}^{\alpha_0-1} E_{it}^{\alpha_1}\right], \end{aligned} \tag{2}$$

where knowledge achieved equals $(1-\delta_i)K_{i,t-1}$ if the student invests zero effort. The value-added specification takes the previous year's knowledge, $K_{i,t-1}$, as a sufficient statistic for student endowments and effort from all previous years.[33] This class of functions can also be written in a cumulative form in which end-of-year knowledge depends on all past inputs and endowments.[34]

3.2 The Student's Decision Problem

Consider a risk-neutral student i whose utility from knowledge net of effort cost is given by:

$$U_i(E_i) = \pi_i K_i + \sum_l \theta_l \mathbf{1}\{K_i + \epsilon_i^S \geq T_l\} - \frac{c}{p}(E_i)^p,$$

where $S_i = K_i + \epsilon_i^S$ is a test score that measures end-of-year knowledge with a normally distributed error term; T_l is the cutoff for letter grade l; θ_l is the student's utility benefit from letter

[32] The regression rate can be thought of as a "depreciation rate" in a loose sense, capturing both how one might forget knowledge if one does not practice it and how applicable one's knowledge from the previous year might be to learning this year's content.

[33] As noted by Todd and Wolpin (2018), this tradition dates back to at least Ben-Porath (1967) who, in eqs. (2) and (4), assumes $\dot{K}_t = \beta_0(s_t K_t)^{\beta_1} D_t^{\beta_2} - \delta K_t$, where $\dot{K}_t$ represents the time (year) derivative of knowledge; s_t, the fraction of time spent learning; D_g, other inputs. Setting $\beta_0 D_t^{\beta_2} \equiv \tau A_{jt}^{\gamma_0} R_{jt}^{\gamma_1}$ and $s_t^{\beta_1} K_t^{\beta_1-1} \equiv K_{i,t-1}^{\alpha_0-1} E_{it}^{\alpha_1}$ leads to the equivalence between his model and this model.

[34] That is, $K_{it} = f(\boldsymbol{\eta}_i)\prod_{s=1}^{t}(1-\delta)\left[1 + \left(\frac{1}{1-\delta}\tau A_{js}^{\gamma_0} R_{js}^{\gamma_1}\right) K_{i,s-1}^{\alpha_0-1} E_{is}^{\alpha_1}\right]$, where $\boldsymbol{\eta}_i$ represents a vector of student endowments at time 0, and $f(\boldsymbol{\eta}_i)(1-\delta)^t$ what the student's knowledge would be at the end of year t if she were to invest zero effort each year.

grade l; π_i is the student's marginal utility of knowledge net of the certification value. I now focus my analysis on a single year, and drop the time subscript for notational convenience ($t \equiv 1$).

The student's decision problem is to maximize her expected utility,

$$\mathbf{E}\,U_i(E_i) = \pi_i K_i + \sum_l \theta_l \Phi\left(\frac{K_i - T_l}{\sigma^s}\right) - \frac{c_i}{p}(E_i)^p, \tag{3}$$

where $\Phi(\cdot)$ represents the cumulative distribution function (CDF) of a standard normal deviate.[35] Note that, because eq. (2) implies a one-to-one mapping between E_i and K_i, eq. (3) is invariant whether the student maximizes it with respect to E_i or K_i.

The decision problem with respect to K_i implies the first-order condition:

$$\underbrace{\pi_i + \sum_l \frac{\theta_l}{\sigma^S}\phi\left(\frac{K_i^* - T_l}{\sigma^S}\right)}_{\substack{\text{marginal benefit of} \\ \text{knowledge, } MB(K^*)}} = \underbrace{a(K_i^* - b)^\lambda}_{\substack{\text{marginal cost of} \\ \text{knowledge, } MC(K^*)}}, \tag{4}$$

where $\phi(\cdot)$ represents the probability density function (PDF) of a standard normal deviate. Note that, for simplicity of notation, I have let $a \equiv \frac{c_i}{\alpha_1}\left[\left(\tau A_j^{\gamma_0} R_j^{\gamma_1}\right) K_{i0}^{\alpha_0}\right]^{-\frac{p}{\alpha_1}}$, a parameter governing the scale of the marginal-knowledge-cost curve; $b \equiv (1-\delta_i)K_{i0}$, a parameter governing the horizontal intercept of the marginal-cost curve; and $\lambda \equiv \frac{p-\alpha_1}{\alpha_1}$, a parameter governing the curvature of the marginal-cost curve.[36]

Going forward, I reduce the dimension of letter grades to one, for two reasons: (1) getting the lowest passing grade is one that matters for eligibility to training in higher-level technical and science-related subjects in this setting; (2) in my sample, there are only a very small mass of students for whom the higher-level cutoffs matter. In order to reflect the fact that the presence of higher letter grades may nevertheless act as continued motivators for these top students, I empirically consider a promotional chance value of the form: $\frac{\theta}{\sigma^S}\phi\left(\frac{\min\{K_i - T, 0\}}{\sigma^S}\right)$.

[35] By the 2-fold rotational symmetry of the standard normal CDF, $\Pr\{\epsilon_i^S \geq T - K_i\} = 1 - \Phi\left(\frac{T-K_i}{\sigma^s}\right) = \Phi\left(\frac{K_i - T}{\sigma^s}\right)$.

[36] Intuitively, the higher the productivity of student effort, the lower the scale parameter; the higher the output elasticity of student effort, the higher the curvature. The intercept represents the level of knowledge in the case of zero student effort.

As seen, the left-hand side of eq. (4) is a constant plus bell curves (that are increasing with K_i), and the right-hand side is a power curve. The nonlinear marginal-benefit structure gives students who are closer to the letter-grade cutoff (T) higher motivational push than it gives students who are farther away from the threshold. Conversely, the structure demotivates students for whom the cutoff is set too far out of reach; these students would begin the year seeing little reason to invest effort in performance.

Finally, I specify minimum productivity thresholds as introduced in the previous section: $\underline{\pi}$ and $\underline{K}_0$. That is,

$$E^*_{PT} = \begin{cases} E^*_{CD} & \text{if } \pi^*_{i,CD} \geq \underline{\pi} \text{ and } K_{0i} \geq \underline{K}_0 \\ 0 & \text{otherwise} \end{cases},$$

where $\pi^*_{i,CD} = MB(K^*_{CD})$, the marginal benefit from the benchmark case. Note that PT can be motivated from fixed costs of the form $g_i \, \mathbf{1}\{E > 0\}$.

I argue that the thresholds provide a parsimonious way of capturing heterogeneous learning curves. In particular, $\underline{K}_0$ may be viewed as a basic entry requirement to making any improvement in the curriculum from the minimum score. Meanwhile, $\underline{\pi}$ may be viewed as entry requirements that are progressively higher for better performing students, governed by the condition $\mathbf{E}\,U(\cdot \mid \pi_i, K_{0i}) > \mathbf{E}\,U(\cdot \mid \underline{\pi}, K_{0i})$.[37] This entry requirement may be motivated by a fixed cost of the form $g_i \, \mathbf{1}\{E_i > 0\}$, where $g_i = \max\{\, \mathbf{E}\,U(\cdot \mid \underline{\pi}, K_{0i}) - \mathbf{E}\,U(E_i = 0 \mid \pi, K_{0i}), 0 \,\}$. Consider an analogy to students learning the basics of algebra. $K_{0i} < \underline{K}_0$ may be likened to students knowing only addition and subtraction but not multiplication: these students cannot obtain any meaningful score in this curriculum. $K_{0i} > \underline{K}_0, \pi^*_i < \underline{\pi}$ may be likened to students knowing multiplication but still needing to incur some fixed costs of effort to digest more complicated concepts in the curriculum, such as quadratic equations, in order to acquire positive value added.

I discuss empirical identification of the model in section 4.5 and appendix C.

[37] Clearly, $\mathbf{E}\,U(\cdot \mid \underline{\pi}, K_{0i})$ is increasing in K_{0i}.

3.3 Mapping to Treatment Effects

The Incentive Only (G1) group can be thought of as receiving a shock to the net-utility parameter, $\pi_i \to \pi_i + \pi^{\$}\nu$, where ν represents the dollar amount of incentive per unit knowledge, and $\pi^{\$}$ represents utils per dollar.[38] The Technology Only (G2) group can be thought of as receiving a shock both to the technology parameter and to the minimum-preparedness-threshold parameter: $\tau^{\text{cons}} \to \tau^{\text{cons}} + \Delta\tau^{\text{SMR}}$, and $\underline{K}_0^{\text{cons}} \to \underline{K}_0^{\text{cons}} + \Delta\underline{K}_0^{\text{SMR}}$, where "cons" stands for the ("constant") status-quo level of technology, and "SMR" stands for the paper's evaluated program. This latter assumption is to reflect the fact that providing more understandable books, such as bilingual books, may reduce the threshold-level preparation required for learning the curriculum (Kremer, Miguel and Thornton, 2009). The Both (G3) group can be thought of as receiving positive shocks to all three parameters, $(\pi_i, \tau^{\text{cons}}, \underline{K}_0^{\text{cons}}) \to (\pi_i + \pi_i + \pi^{\$}\nu, \tau^{\text{cons}} + \Delta\tau^{\text{SMR}}, \underline{K}_0^{\text{cons}} + \Delta\underline{K}_0^{\text{SMR}})$.

3.4 Valuations and Welfare

This framework admits valuation of utils in dollar terms based on revealed-preference theory. Students value certification at $\frac{\theta}{\pi^{\$}}$ dollars. The marginal value of knowledge net of the certification value is $\frac{\pi_i}{\pi^{\$}}$ dollars. These valuations jointly characterize students' willingness to pay for knowledge. Similarly, I can assess the value of each student's marginal hour of time. I can also compare and contrast the welfare increases caused by the treatment interventions, and evaluate the welfare increase caused by the interaction effect of incentive and technology. Total revealed-preferred student welfare is given by $\int \frac{1}{\pi^{\$}} \mathbf{E}\, U_i \, \mathrm{d}i$.[39]

[38] This assumes that incentives do not crowd out intrinsic motivation, often shown to be the case in economic settings (Prendergast, 2011).

[39] Such a multiplicative translations into welfare terms are commonly seen in public economics. Carleton et al. (2018), for example, multiply value-of-statistical-life estimates with mortality effects of adapting to climate change to value mortality-related health benefits of adaptation.

3.5 Counterfactual Scenarios and Optimal Certification Cutoff

In defining policy objectives, communities may have different preferences as to how to assign relative weights over different educational outcomes. A community may be interested in maximizing aggregate knowledge, $\int K_i \, \mathrm{d}i$. A community may be interested in maximizing aggregate effort, $\int E_i \, \mathrm{d}i$.[40] A community may be interested in maximizing private welfare of students, $\int \mathbf{E} U_i \, \mathrm{d}i$.[41] In considering counterfactual certification policies below, I assume that the community's objective function is to maximize effort; analogous considerations could be made for alternative objectives.[42]

Recall that two key ideas of this paper are that (1) the extent to which students deem certification "attainable" may be a strong motivator of student performance, and that (2) educational technologies (or learning materials) may be strong complements to motivation. To the extent that (1) is important, lowering the cutoff may be as powerful a motivator as providing cash incentives for test scores, especially for students toward the middle of the rising portion of the marginal benefits curve. Yet, to the extent that (2) is important, a policy of lowering the cutoff (or otherwise making the certification test more accessible) may not lead to significant increases in effort, if students are also commonly lacking in the means by which to practice learning. Hence, a community's optimal certification policy may depend crucially on the level of educational technology commonly supplied in the community. That is, it may be important to consider,

$$\left\{T^*, \tau^*\right\} = \underset{\{T,\tau\}}{\arg\max} \int \mathbf{E} E_i^*(T,\tau) \, \mathrm{d}i - p_\tau \tau, \tag{5}$$

[40] There may be value apart from increasing knowledge in curricular content, for example, in getting students to practice following directions or collaborating with each other. Although there is a one-to-one mapping between knowledge and effort in this framework, individuals differ in how their effort maps to knowledge, and effort is also a convex function of knowledge. This implies that maximizing average knowledge would mean prioritizing higher-performing students, and maximizing average effort would mean prioritizing lower-performing students.

[41] If social returns to outcomes such as knowledge and effort exceed private returns, however, a community may prioritize these other outcomes over private welfare alone. For example, there may exist knowledge spillovers, and student discount rates may be lower than the community-wide discount rate.

[42] A related angle is to consider to which objective the government's current certification policy maps the closest.

where p_τ stands for the cost of supplying technology τ.

The assumption that θ remains constant in counterfactual scenarios warrants a discussion. I justify it based on two contextual reasons. First, in this setting, the margin of additional passing in mathematics is to come from those already expected to secure seats at the A Level by scoring sufficiently well on O Level subjects other than mathematics. Second, field interviews suggest that mathematics education at the A Level involves minimal teaching and mostly self- and group-study given printed materials, further assuaging congestion concerns.

A potentially meaningful corollary to this analysis is that a community may not be able to reap significant benefits from policies designed to make aspects of a certification process (content/curriculum/cutoff) more "accessible," without also commonly supplying appropriate technologies to practice with (and vice versa). Across educational settings, there may be systemic barriers that make it difficult for communities to consider pulling both policy levers at the same time. Yet, a simple change in perspective might just be what is needed to make progress from the status quo: holding technology fixed, a high cutoff could seem optimal because lowering it might just mean passing more students who have not learned much; holding a high cutoff fixed, supplying a more efficient learning technology alone might not produce much reaction among students, leading to community inaction. I check the empirical magnitudes of these implications.

4 Structural Estimation

I rely on a simulated-maximum-likelihood approach, estimating three categories of structural parameters:

(1) the parameters of the production and utility functions, as outlined above;

(2) the parameters of a latent-factor model, specifying how endowments are determined by exogenous initial conditions, some of which I assume are measured in survey responses and others (classroom- and individual-level unobservables) I draw from simulations;

(3) the parameters of a measurement-error model, which specifies how knowledge, effort and endowments are measured.

I first adapt the assumptions and associated identification arguments of the likelihood approach largely from Todd and Wolpin (2018). The assumptions provide a high degree of flexibility in accounting for uncertainties that may be inherent in student test-score and survey data.

I then contribute a novel identification result involving nonlinear (cutoff-based) returns and a computational technique of searching for solutions in the intermediate-output space (K space) rather than in the input space (E space). I also additionally model selection explicitly to account for attrition, and truncation explicitly to account for limited ranges that test scores and other variables can take in the observed data.

I calculate the joint likelihood of observing the measurement outcomes of each classroom j based on the joint measurement-error distribution, taking the average across simulation draws.[43] I integrate the likelihoods of each classroom over the whole sample of classrooms, and then maximize the sample joint likelihood over the parameter vector space.

4.1 Latent Factor Structure

I assume student endowments, $\boldsymbol{\eta}_{ij} = \{K_{i0}, \pi_i, A_j, R_j\}$, are latent factors measured with error.[44] Each factor $\eta_{ij} \in \boldsymbol{\eta}_{ij}$ depends on a set of exogenous initial conditions, $\boldsymbol{X}^{\eta}_{ij}$, and unobserved classroom- and individual-difference components, μ^{η}_{j} and ω^{η}_{ij}:

$$K_{ij0} = \boldsymbol{X}^{K_0}_{ij}\boldsymbol{\beta}^{K_0} + \mu^{K_0}_{j} + \omega^{K_0}_{ij}, \tag{6}$$

$$\pi_{ij} = \boldsymbol{X}^{\pi}_{ij}\boldsymbol{\beta}^{\pi} + \mu^{\pi}_{j} + \omega^{\pi}_{ij} + \pi^{\$} \times (G^1_i + G^3_i), \tag{7}$$

$$A_j = \boldsymbol{X}^{A}_{j}\boldsymbol{\beta}^{A} + \mu^{A}_{j}, \tag{8}$$

$$R_j = \boldsymbol{X}^{R}_{j}\boldsymbol{\beta}^{R} + \mu^{R}_{j}. \tag{9}$$

[43] Classroom j and teacher j are equivalent in this setting.

[44] Assuming no student gets negative utility from knowledge, I impose that the latent factors are bounded from below by zero.

All difference components are assumed to be mean zero, orthogonal to each other and orthogonal to observed characteristics, for a given endowment. Across endowments, the difference components may be freely correlated.[45] Note that in eq. (7), $\pi^{\$}$ identifies the per-mark utility benefit of the experimental piece-rate incentive.

4.2 Truncation and Selection Structures

In order to account for test-score truncation, I modify eq. (3) in the following way.

$$\mathbf{E}\,U_i(E_i) = \pi^{\$}(K_i - S^{\min})\Phi\left(\frac{K_i - S^{\min}}{\sigma^{\mathrm{s}}}\right) + \pi^{\$}\sigma^{\mathrm{s}}\phi\left(\frac{K_i - S^{\min}}{\sigma^{\mathrm{s}}}\right) \tag{10}$$

$$+ \pi_i K_i + \theta\,\Phi\left(\frac{K_i - T}{\sigma^{s}}\right) - \frac{a}{\lambda+1}(K_i - b)^{\lambda+1}, \tag{11}$$

where the first of the two added terms (in the first line) describes utility benefit of SMR incentives scaled by $\Pr\{S_i > S^{\min}\}$, and the second term describes the expected truncation bonus.[46]

The modification of eq. (4) follows accordingly:

$$\pi_i + \pi^{\$}\Phi\left(\frac{K_i - S^{\min}}{\sigma^{\mathrm{s}}}\right) + \frac{\theta}{\sigma^{S}}\phi\left(\frac{K_i^* - T}{\sigma^{S}}\right) = a(K_i^* - b)^{\lambda}. \tag{12}$$

Selection is modeled in the following manner. Students are assumed to be required to pay a random test attendance cost:

$$\zeta_{ij} = \boldsymbol{X}_{ij}^{\zeta}\boldsymbol{\beta}^{\zeta} + \epsilon_{ij}^{\zeta}, \tag{13}$$

where ϵ_{ij}^{ζ} is assumed to be an independent standard normal. If a student does not attend, the student reduces learning by ι, modeled as a portion of effort. On test day, student attends if

$$[\mathbf{E}\,U_i(E_i^*) - \mathbf{E}\,U_i(E_i^* - \iota)] \times \beta^{\mathrm{Udiff}} > \zeta_i, \tag{14}$$

[45] In implementation, I do not allow for difference components at some levels to conserve on parameters.

[46] If $S_i < S^{\min}$, the student gets $\pi_i K_i$. If $S_i > S^{\min}$, the student gets $\pi_i K_i + \pi^{\$}(S_i - S^{\min})$. Taking expectations gives the terms.

and avoids the test otherwise. Therefore, the likelihood of observing an absent student's observation is given by $\Phi(\zeta_i - [\mathbf{E}\,U_i(E_i^*) - \mathbf{E}\,U_i(E_i^* - \iota)] \times \beta^{\text{Udiff}})$, and the likelihood of observing a present student's observation is scaled by $1 - \Phi(\zeta_i - [\mathbf{E}\,U_i(E_i^*) - \mathbf{E}\,U_i(E_i^* - \iota)] \times \beta^{\text{Udiff}})$.

4.3 Measurement Structure

Given $m = 1, \ldots, M^{\eta}$ measures for each latent factor η, I assume measurement equations given by,

$$\boldsymbol{\eta}^m = \beta_0^{\boldsymbol{\eta},m} + \beta_1^{\boldsymbol{\eta},m} \boldsymbol{\eta}_{ij0} + \epsilon_{ij}^{\boldsymbol{\eta},m}, \quad m = 1, \ldots, M^{\boldsymbol{\eta}}, \tag{15}$$

where $\eta \in \boldsymbol{\eta} = \{K_{i0}, \pi_i, A_j, R_j\}$ as before. All measurement errors are assumed to be uncorrelated with all of the latent variables (both observed and unobserved components) and with each other.

I also treat student effort as a latent variable measured with error.[47] Given M^E measures of student effort, the effort measurement equation is given by,

$$E_{ij}^m = \beta_0^{E,m} + \beta_1^{E,m} E_{ij} + \epsilon_{ij}^{E,m}, \quad m = 1, \ldots, M^E. \tag{16}$$

Students determine their levels of effort by solving the knowledge decision problem (eq. (3)), which is fully determined by the latent endowments and fixed production-function parameters.

I additionally estimate a location parameter of latent student effort, one that may be interpreted as reporting bias. That is, I estimate $\underline{E}$ in the relationship,

$$E_{ij}^{\text{latent}} = \underline{E} + E_{ij}^* + \epsilon_{ij}^{E^{\text{latent}}}. \tag{17}$$

Several interpretations may apply to $\underline{E}$. If $\underline{E}$ is positive, it may mean that students tend to

[47] This methodology reduces the dimension of the effective input space to one (Cunha and Heckman, 2008). In table 4 and table 6, hours of mathematics study in response to the treatments move similarly to how test scores move, while other candidate proxies for effort do not.

overstate their effort, or that this part of effort is "unproductive": for example, students report nontrivial positive numbers in response to the question, "what percentage of this mathematics study time were you paying attention and not copying from friends?"

End-of-year knowledge is a latent variable measured by end-of-year test marks, S_{ij}, with error:

$$S_{ij} = K_{ij} + \epsilon^{S}_{ij}, \tag{18}$$

where K_{ij} is determined by eq. (2).

A measurement outcome consists of (i) test marks, measures of effort level, and measures of initial knowledge and interest levels of each student; and (ii) measures of effort level and ability of teachers. A measurement outcome is conditioned by the absence indicator. I denote the set of measurement outcomes for classroom j as:

$$O^{M}_{j} = \left\{ S_{ij}, E^{\boldsymbol{m}}_{ij}, K^{\boldsymbol{m}}_{ij0}, \pi^{\boldsymbol{m}}_{ij}, A^{\boldsymbol{m}}_{j}, R^{\boldsymbol{m}}_{j}, Absence_{ij} : i = 1, \ldots, N_j \right\}. \tag{19}$$

I additionally denote the set of observable determinants of latent endowments as $O^{X} = \left\{ \boldsymbol{X}^{\boldsymbol{\eta}}_{ij} \right\}$.

I assume that the vector of unobservables and measurement errors are jointly normal. Specifically, let $\boldsymbol{\upsilon}^{\mu} = \left\{ \mu^{\boldsymbol{\eta}}_{j} \right\}$ represent the vector of classroom-level observables; $\boldsymbol{\upsilon}^{\omega} = \left\{ \omega^{K_0}_{ij}, \omega^{\pi}_{ij} \right\}$ the vector of student-level unobservables; and $\boldsymbol{\upsilon}^{\epsilon} = \left\{ \epsilon^{\boldsymbol{\eta},\boldsymbol{m}}_{ij}, \epsilon^{E,\boldsymbol{m}}_{ij}, \epsilon^{S}_{ij} \right\}$ the vector of measurement errors. The unobservables $\boldsymbol{\upsilon}^{\mu}$ and $\boldsymbol{\upsilon}^{\omega}$ have joint distributions F_{μ} and F_{ω}, assumed to be normal with variance-covariance matrices Σ_{μ} and Σ_{ω}. The measurement errors $\boldsymbol{\upsilon}^{\epsilon}$ have a joint distribution F_{ϵ}, assumed to be normal with a diagonal variance-covariance matrix Σ_{ϵ}.

4.4 Likelihood Function

Estimation is carried out by simulated maximum likelihood. The likelihood contribution of classroom j is the joint density of O_j. The estimation routine proceeds as follows:

1. Guess $\{\alpha_0, \alpha_1, \gamma_0, \gamma_1, \boldsymbol{\delta}, \tau^{\text{cons}}, \tau^{\text{SMR}}, \pi^{\$}, \theta, \boldsymbol{\beta}^{\boldsymbol{\eta}}, \Sigma_{\mu}, \Sigma_{\omega}, \Sigma_{\epsilon}, \underline{E}, \underline{\pi}, \underline{K}_0, \underline{K}^{SMR}_0, \iota, B^{\boldsymbol{m}}\}$: a parame-

ter vector where $B^{\boldsymbol{m}}$ denotes the set of $\beta_0^{\boldsymbol{\eta},\boldsymbol{m}}$ and $\beta_1^{\boldsymbol{\eta},\boldsymbol{m}}$ parameters of the measurement error equations.

2. Draw shocks $\boldsymbol{\upsilon}^{\mu}$ and $\boldsymbol{\upsilon}^{\omega}$ for all students $i = 1,\ldots,N_j$ and classrooms $j = 1,\ldots,J$.
3. Given the shocks and observed determinants O^X, compute the values of endowments $\boldsymbol{\eta}_{ij}$.[48]
4. Compute K^* and E^* for all students in all classrooms for all draws $d = 1,\ldots,D$.[49]
5. For each draw d, given the joint measurement-error distribution F^{ϵ}, calculate the joint likelihood of the measurement outcome O_j^M.[50] Denote this joint density as $f_j(d)$.
6. Compute the mean value of the joint density across draws: $\mathcal{L}_j = \frac{1}{D}\sum_d f_j(d)$.
7. Repeat for all $j = 1,\ldots,J$ classrooms. The likelihood of the entire sample is $\prod_{j=1}^{J}\mathcal{L}_j$.
8. Repeat steps 1 through 7, maximizing the sample likelihood over the space of parameter vectors.

4.5 Identification

An innovation in this paper is demonstrating empirical identification even in the case of nonlinear returns to knowledge given the possibility of certification. Apart from the innovation, identification is via Cunha, Heckman and Schennach (2010). I discuss identification separately for the parameters of the production function, and the parameters in the latent-factor and measurement-error equations.

First, suppose I have perfect measurements of A_j, R_j, K_{i0} and E_{ij}. Combining eq. (18) and eq. (2) gives us:

$$S_{ij} = (1-\delta_i)K_{i0} \times \left[1 + \left(\frac{1}{1-\delta}\tau A_j^{\gamma_0} R_j^{\gamma_1}\right) K_{i0}^{(\alpha_0-1)} E_i^{\alpha_1}\right] + \epsilon_{ij}^S.$$

[48] I set to zero any latent factor that is negative.

[49] It is computationally more efficient to search for K^* first and then back out E^* by inverting eq. (2), this paper's technical innovation.

[50] Some survey measures are continuous, some ordered-categorical, some dichotomous, and all bounded. In all cases, I assume a continuous latent measure that underlies the observed measure; I treat the bounded measures as truncated, the dichotomous variables as probits, and ordered-categorical variables as ordered probits.

Given that the only student-level unobservable is the test-score measurement error, which is assumed to be orthogonal to the determinants of K_i, identification of the production function parameters follows immediately from independent variations in the perfect measurements.

I do not assume to have perfect measurements of A_j, R_j, K_{i0} and E_{ij}; however, with multiple measures, this framework folds into a special case of Theorem 2 in Cunha, Heckman and Schennach (2010). The measurement equations of section 4.3, together with the implicit determination of E^* as an argument optimum of the student's decision problem, correspond to the system of nonlinear measurement equations given by (3.7) in Cunha, Heckman and Schennach (2010). Given the orthogonality conditions, the parametric forms of the measurement equations, and distributional assumptions about the measurement errors, I can invoke their theorem to identify the production function parameters.

I can easily identify the parameters in the latent-factor and measurement-error equations, based on the linearity of the equations and orthogonality of the unobserved terms, following Todd and Wolpin (2018). For illustration, consider two measures of a latent factor η, with measurement equations,

$$\eta_{ij}^{m1} = \eta_{ij} + \epsilon^{\eta,m1}, \tag{20}$$

$$\eta_{ij}^{m2} = \beta_0^{\eta,m2} + \beta_1^{\eta,m2}\eta_{ij} + \epsilon^{\eta,m2}, \tag{21}$$

where, without loss of generality, I allow for only a student-level unobservable. Note the normalization $\beta_0^{\eta,m1} = 0$ and $\beta_1^{\eta,m1} = 1$, which establishes the measure $m1$ as the metric of η_{ij}. The latent-factor equation is given by,

$$\eta_{ij} = \beta_0^{\eta} + \beta_1^{\eta} X_{ij}^{\eta} + \omega_{ij}^{\eta}, \tag{22}$$

where, without loss of generality, I assume X_{ij}^{η} is scalar; there is only a student-level unobservable; and all unobservables are orthogonal to each other and to X_{ij}^{η}. Note that β_0^{η} and β_1^{η} are identified upon regressing η_{ij}^{m1} on X_{ij}^{η} after substituting eq. (22) into eq. (20). The factor loading

in the second measurement equation, $\beta_1^{\eta,m2}$, is given by $\text{Cov}\left(\eta_{ij}^{m2}, X_{ij}^{\eta}\right) \Big/ \left[\beta_1^{\eta}\text{Var}\left(X_{ij}^{\eta}\right)\right]$; the location parameter, $\beta_0^{\eta,m2}$, is then identified by passing the line through the means. The variance of the unobservable, $\text{Var}\left(\omega_{ij}^{\eta}\right)$, is derived from the covariance between the two measurements, $\text{Cov}\left(\eta_{ij}^{m1}, \eta_{ij}^{m2}\right)$.[51] The measurement error variances are derived from the variances of the measures. The same argument applies to all other measures.

To complete the identification argument, I address the last remaining step, which is that the nonlinear marginal benefit of knowledge does not admit multiple levels of knowledge that deliver the same utility, an innovation of this paper. It can be seen that in eq. (4), the left-hand side is an exponential curve while the right-hand side is a power curve, whose intersection admits discrete roots. Therefore, while the two curves may cross each other at multiple points, that they do so at equal levels of utility is a measure zero event when the latent factors are being drawn from a continuous distribution, and the student's solution is a.s. unique. See appendix C.

4.6 Empirical Determinants and Measures

For the structural estimation, I estimate the model in year 2, taking year 0 and year 1 responses as predetermined characteristics.[52] The full list of determinants and measures can be seen in table A8 and table A9. The survey questions broadly follow Todd and Wolpin (2018). I estimate two models: the benchmark model without productivity thresholds (model 1), and the proposed model with productivity thresholds (model 2).

Items 20 through 31 list the variable determinants assumed for K_0. These include the female indicator; year 0 math score; age; FTNA nonmath average; and parental education (high school or above).

Items 33 through 36 list the variable determinants assumed for π_0. These include female indicator; parental education; FTNA nonmath average; and whether the students' reported desired job was a STEM occupation.

[51] $\text{Cov}\left(\eta_{ij}^{m1}, \eta_{ij}^{m2}\right) = \beta_1^{\eta,m2}\text{Var}\left(\eta_{ij}\right) = \left(\beta_1^{\eta}\right)^2\text{Var}\left(X_{ij}^{\eta}\right) + \text{Var}\left(\omega_{ij}^{\eta}\right)$.

[52] A key condition for identification—two or more independent measures of previous year test scores—is not met in years 1 or 3.

Items 38 through 45 list the variable determinants assumed for A_j. These include being a full-time math teacher (as opposed to part-time or substitute teacher whose specialty is another subject); has bachelor's degree; number of years taught math; number of years taught math squared.

Items 47 through 51 list the variable determinants assumed for R_j. These include the total number of teaching hours per week; being a full-time math teacher; has bachelor's degree; number of years taught math; number of years taught math squared.

Items 53 through 57 list the variable determinants assumed for attendance probability. These include the utility difference between attending and skipping school; parental education (high school or above); female; age; parental education and commute distance (which serves as an instrument for selection correction).

Items 68 through 95 indicate measures used for the latent factors. K_0 has two measures: year 1 math score; and FTNA math grade. π_0 has two measures: year 0 reported hours of non-math study; and year 0 reported degree to which student "likes math"—a categorical ordered from 1 ("never") to 4 ("always"). A_j has two measures: the proportion of students in class reporting that the teacher "always knows the subject"; the proportion reporting that the teacher "always has control of class." R_j has two measures: the proportion of students in class reporting that the teacher "always cares that the student learn"; the proportion reporting that the teacher "always cares that the students pay attention." E has three measures: year 2 reported hours of math study; year 2 reported percentage of attention paid on homework; and year 2 reported degree of effort expended on the test—a categorical ordered from 1 ("low") to 4 ("high").

All test marks are converted to scaled scores with mean 500 and standard deviation 100, following the convention of many previous empirical works, as well as international educational authorities such as Program for International Student Assessment (PISA) and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).

5 Estimation and Simulation Results

5.1 Parameter Estimates across Models

Table 7 reports selected parameters and simulated expected treatment effects, comparing them against the reduced-form treatment effects.[53] Table 8 reports valuation estimates. Model 1 refers to the benchmark model. Model 2 refers to the model with productivity thresholds. Figure 5 plots heterogeneous model-simulated treatment effects side-by-side, in the same manner as in figs. 4b and 4c, to highlight comparisons with reduced-form effects and also maximize the contrast of performance between models.

Indeed, fig. 5 highlights the strength of the productivity-thresholds model. The figure juxtaposes difference-in-means coefficients simulated based on one arbitrary simulation draw from each of the two structurally estimated models of student learning, thus allowing for individual-level unobserved factor heterogeneity. It can be seen that fig. 5b, which plots the benchmark model's value-added interaction effects, displays no discernible hump-shaped pattern. In contrast, fig. 5d generates the hump-shaped pattern seen in the reduced-form effects of fig. 4c, with almost identical levels of confidence.[54] In addition, fig. 5c matches the two other patterns alluded to in section 2.3 that fig. 5a does not: how (1) the bottom quintile of students seem to respond to money alone; and (2) the top quintile of students seem to respond to technology alone. In light of the pattern seen in the value-added interaction effects, the gross differences suggest that (1) is in large part due to sampling variation; and (2) is in large part due to the top quintile of students entering the grade both more motivated (by starting closer to the promotional cutoff) and better prepared (more likely to meet the minimum preparedness threshold) than the rest.

Figure 6 shows how the productivity-thresholds model produces the effects across different scenarios of structural component inclusion. In scenario 1 (S1), "NO promotion concern"

[53] The entire sets of parameter estimates and their standard errors are reported in tables A8 and A9.

[54] Figures 5b and 5d display differences in "value-added" (i.e., $K_i - (1-\delta_i)K_{i0}$) to make salient the interested mechanism of effective effort.

means $\theta = 0$; "NO knowledge threshold" means $\underline{K}_0 = 0$; "NO interest threshold" means $\underline{\pi} = 0$. Results over-predict the technology-only (G2) effect, and cannot generate the interaction effects. Ensuing scenarios are based on adding different combinations of structural components as labeled. In scenario 2, adding promotion concerns alone to scenario 1 generates negligible difference from scenario 1. In scneario 3, adding the knowledge threshold (without the interest threshold) exacerbates the over-prediction of the G2 effect. In scenario 4, it can be seen that adding the interest threshold alone begins to generate the observed shape of effects in the data. Scenario 5 shows the preferred combination. In scenario 6, it can be seen that removing the promotion concern while leaving the threshold concerns in erases the effects seen on the top students.

In table 7, panel A reports estimated production function parameters. α_0, the output elasticity of knowledge endowment, is large; that the point estimate of this parameter is greater than 1 was also seen in Todd and Wolpin (2018), though not significantly greater than 1. In contrast, α_1, the output elasticity of effort, is estimated to be approximately one seventh of α_0. Teacher ability, measured at least within the range of variation observed in the data, is estimated to matter little in model 1, and not matter at all in model 2; teacher effort is also predicted to matter negligibly within the range of variation observed in this setting.

The baseline "regression" rate, $\delta^{\text{cons}} = 0.148$, is estimated to be large in model 1; small in model 2.[55] In both models, the depreciation rate falls as the pre-score rises: for each 100 rise in scaled score (i.e., 1 standard deviation rise) above the minimum score, model 1 predicts a fall in the depreciation rate of 0.012; model 2 predicts a fall of 0.033.[56] Model 1 simulations suggest that control-group students are gaining an average knowledge of 0.65σ in one year, and even those from the bottom pretest quintile (those getting average percentage marks of three out of hundred) are gaining 0.53σ of knowledge in one year—almost three and a half years' worth

[55] Again, this is "depreciation" in a loose sense: capturing both how the total knowledge level might regress if not put to use and how difficult it might be to apply knowledge from the previous year to performance this year.

[56] For top students (e.g. five standard deviations above the minimum score), this means that a standard deviation gain in knowledge endowment halves in less than four years, suggesting that even investments at younger ages would not remain effective if only temporarily applied in this type of curriculum.

of progress according to nationally-normed scales in other contexts.[57] In contrast, model 2 simulations suggest that control-group students are gaining 0.07σ of knowledge throughout the year and bottom students are gaining 0.0003σ of knowledge, perhaps offering greater fidelity to observed empirical realities: students from the bottom quintile score around 3 marks out of 100 in expectation, and less than 0.003% pass the final O Level mathematics examination.

The SMR technology is seen to improve value-added factor productivity (τ) by 15% in model 1, and 35% in model 2.

Model 1 ascribes a lower marginal utility of income (4.79) than model 2 (14.1). As can be seen in the second row of table 8, this leads model 1 to ascribe a value of \$13.2 per 10 scaled-score units, while model 2 to ascribe a value of \$4.20. π_{scale}, a scale parameter included to potentially control for scale bias in reported measures for π_i, varies within 10% of each other across models. The value of promotion is estimated to be much larger in model 1 (\$289) than model 2 (\$26.6). Model 1 ascribes a lower cost elasticity of effort, lower location bias (or amount of unproductive effort), and lower noise, consistently predicting higher productivity and efficiency of the education system than model 1.

At the bottom of panel A, I report the proportion of students model 2 attributes as being either "disinterested" ($\pi_i < \underline{\pi}$); "lost" ($K_{0i} < \underline{K}_0$); both disinterested and lost; and neither. Only 12.8% of students are seen to be meaningfully accumulating knowledge in model 2, whereas in model 1, by construction, every student is equating her marginal cost of effort to a highly estimated marginal benefit of knowledge.

In panel B, column (Data) reproduces estimates from column (6) of table 3, and the bottom row of table 5. Columns (1) and (2) report differences in simulated means across treatment groups, leaving the original balance of student characteristics across treatment groups intact. The means are obtained across 1,000 simulation draws; hence, the differences are "model-expected" treatment effects net of uncertainties from measurement errors as well as

[57] An often-targeted benchmark of success in educational interventions is 0.1σ. In an analysis of nationally-normed tests in the US (originally adapted from Hill et al. (2008)), Lipsey et al. (2012) note that students typically gained an average achievement of 0.16σ across reading, math, science and social studies over their 11th-grade academic year.

unobserved school-level and individual-level effects (but NOT net of effects of sampling variation such as demographic characteristics). Note that the reduced-form treatment effects in column (1) were NOT targeted during estimation, and as such stand as yardsticks of model performance.[58]

Columns (1') and (2') report differences in simulated means across treatments, assuming that the whole sample gets shocked with each treatment. Although both models underhit the magnitudes of treatment effects seen in the reduced-form data, model 2 consistently generates the patterns of effects better than model 1. In fact, in column (1'), the interaction effects are seen to be off by two to three orders of magnitude; essentially, model 1 does not allow for any meaningful interaction effect to occur. In contrast, though restricted in magnitudes, model 2 generates the observed pattern of hump-shaped dynamics in the interaction effects with respect to the students' pretest percentile.

In panel C of table 7, columns (2) and (2') report differences in simulated proportions of students who are gaining knowledge—that is, those who are neither disinterested nor lost—across treatments, providing a clear intuition for how the treatment effect and hump-shaped patterns were generated. In particular, the both group is seen to have raised the proportion of students meaningfully accumulating by 20 p.p., generating the patterns of average and heterogeneous interaction effects.

Panel A of table 8 reports model-based valuation estimates for each considered educational good. At the bottom of each horizontal panel, it can be seen that model 1 predicts net welfare loss from the interaction effect; that is, the benchmark model suggests that it is more cost effective to just give books than to give both books and incentivize students, while model 2 consistently predicts a welfare gain, intuitively from students crossing the productivity thresholds. Welfare impacts are estimated to be consistently higher in model 1 because model 1 ascribes a lower, less precise marginal utility of income. Model 1 is flexible enough to attribute the differences in outcomes between the both (G3) group and the technology-only (G2) group to a high

[58] 1 normalized-test-score unit equals 100 scaled-score units.

marginal utility of income, whereas model 2 is flexible enough to do so.

Column labeled "(2; entry cost)" shows average valuations of the treatment interventions, interpreting minimum interest thresholds as stemming from entry costs of learning new material on the curriculum. Strikingly, valuation estimates are dramatically reduced, and none of the estimated private welfare impacts are greater than the cost of supplying the interventions through the program. This result illustrates the importance of taking into account potentially steep entry costs of climbing the initial portion of individual learning curves. This is despite the fact that the estimated cost of asking students to cross the knowledge threshold on their own (approximately $67, unreported in the table) is estimated to be an order of magnitude higher than the cost of providing the experimental technologies through the program ($6.13). This is because the interest threshold remains an overwhelming barrier, as seen in fig. 6.

The top of panel B of table 8 reports mean effort cost estimates per 0.01σ, 0.10σ, and 0.2σ of performance gain; what happens to these costs when provided the technology (G2) shock; and mean program costs of the treatments. The effort-cost estimates reveal a highly convex and inelastic structure in both models, but because everyone is pushing themselves to the inelastic margin in model 1, the cost estimates in model 1 are much higher.

5.2 Counterfactual Simulation Results

Figure 7 plots simulated outcomes from counterfactual promotion cutoffs.

The promotion cutoffs are minimum absolute marks of test scores a student must achieve in order to have the option of continuing in STEM training beyond the 11th grade. Plots indicate model-simulated outcomes, *conditional on distributing the experimental technology nationally*, from a counterfactual policy of lowering the current promotion cutoff (indicated in red and also labeled "T: 29.50") to a hypothetical cutoff (indicated in green). The hypothetical cutoff is set to double the number of students endogenously passing in equilibrium. Plotted are nationally allocated estimates, weighting each experimental-sample observation by the ratio between a Weibull density fitted to FTNA grade distribution and the density of the experimen-

tal sample's year 1 marks. "Share Accumulating Knowledge" refers to the proportion of students who are neither disinterested nor lost, thus meaningfully accumulating knowledge during the year.

As seen in the "Promotion Rate" plot, lowering the absolute mark from 29.5 to 19.5 (model 1) and to 20.5 (model 2) would double the endogenous proportion of students passing the test; the mark is lower for model 1 because there is less scope of endogenous response in model 1 where incentives have limited effects.

As seen in the "Knowledge" plot, in model 1, such a policy change would have a negligible, 0.004σ effect on knowledge, despite model 1 attributing a much greater monetary value to passing. This is because model 1 predicts that all students are pushing themselves to a highly inelastic portion of the knowledge-cost curve and are leaving no effort on the table. Model 2, on the other hand, predicts that such a policy change would have a modest but meaningful effect on endogenous response of student knowledge, on the order of 11 scaled-score points (0.11σ), by raising the share of students meaningfully accumulating knowledge by 20 p.p.

6 Conclusion

This paper presented an economic framework for analyzing how students' own agency affects educational outcomes, and used field-experimental treatments to estimate the framework-based model. The estimates were used to evaluate the performance and counterfactual policies of the secondary mathematics education system in Tanzania, where pass rates have remained below 20 percent nationally.

The field experiment tested to what extent the lack of student interest, the lack of basic technologies, or the lack of both thereof might explain the system's performance. I find that neither performance-based incentives, nor free solar-energy access, nor bilingual textbooks, nor videos by themselves could lead to a meaningful performance gain. Providing all these inputs together, on the other hand, showed strong and significant impacts on year-to-year in-

centivized mock tests over a period of three years (0.3σ). Treatment complementarity between incentives and technologies was particularly strong for the scores of students just below the top 20 percent, revealing an inverse-U-shaped relationship between entering grade-level performance and treatment complementarity. This result suggested that a substantial number of students at the middle to upper-middle range of the baseline performance distribution lacked both interest and requisite learning support.

I then presented structural estimates of the students' cost-benefit considerations. In the model, heterogeneous students balance the cost of effort against the benefit of knowledge, given (i) probabilistic proximity to the promotional cutoff, and (ii) preference for knowledge net of the proximity value. On the cost side, I tested two simple specifications: model 1, which has a convex variable cost; and model 2, which has not only the variable cost but also entry costs—minimum interest and minimum preparedness requirements—that may prevent some students from even getting started with new material. Comparisons of model-generated treatment effects against reduced-form treatment effects (untargeted moments) show that model 2 generates key treatment effect patterns, while model 1 cannot, attributing a higher degree of realism to model 2.

Based on model 2, simulations of a counterfactual policy of *both* providing the experimental technologies to students *and* lowering the promotional cutoff suggest that this policy would lead to a modest but meaningful gain in endogenous responses of student knowledge, by reducing the share of students who are giving up on learning new topics from 79% to 52%. However, although the cost of providing technologies through the program is lower than the estimated cost of asking students to attain the equivalent preparation on their own, taking into account the effort cost associated with the "entry" portion of individual learning curves reduces conventional estimates of the interventions' revealed-preference-based welfare impacts by more than two thirds. By explaining treatment complementarities, the proposed analytical framework extends models of classroom learning and welfare implications of policies outlined in the previous literature to reflect a higher degree of realism about developing community contexts.

References

Agostinelli, Francesco, and Matthew Wiswall. 2016. "Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development." National Bureau of Economic Research Working Paper 22441.

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): 1481–1495.

Angrist, Joshua, and Victor Lavy. 2002. "New Evidence on Classroom Computers and Pupil Learning*." *The Economic Journal*, 112(482): 735–765.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton. 2016. "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India." National Bureau of Economic Research Working Paper 22746.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics*, 122(3): 1235–1264.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics*, 3(2): 167–195.

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin. 2015. "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools." *Journal of Political Economy*, 123(2): 325–364.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75(4): 352–365.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika*, 93(3): 491–507.

Bergquist, Lauren Falcao, and Michael Dinerstein. 2019. "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya." University of Chicago Working Paper.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249–275.

Bloom, Howard S. 2005. *Learning More from Social Experiments: Evolving Analytic Approaches*. Russell Sage Foundation. Google-Books-ID: MuSFAwAAQBAJ.

Carleton, Tamma, Michael S. Delgado, Michael Greenstone, Trevor Houser, Solomon M. Hsiang,

Andrew Hultgren, Amir Jina, Robert E. Kopp, Kelly McCusker, Ishan Nath, James Rising, Hee Kwon Seo, Justin Simcock, Arvid Viaene, Jiacan Yuan, and Alice Tianbo Zhang. 2018. "Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits."

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy*, 118(3): 409–432.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Cristia, Julian, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Severín. 2017. "Technology and Child Development: Evidence from the One Laptop per Child Program." *American Economic Journal: Applied Economics*, 9(3): 295–320.

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources*, 43(4): 738–782.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78(3): 883–931.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics*, 1(3): 111–134.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*, 101(5): 1739–1774.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Chapter 61 Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*. Vol. 4, , ed. T. Paul Schultz and John A. Strauss, 3895–3962. Elsevier.

Enders, Craig K. 2010. *Applied Missing Data Analysis*. Guilford Press. Google-Books-ID: MN8ruJd2tvgC.

Foster, Andrew D, and Mark R Rosenzweig. 2017. "Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size." National Bureau of Economic Research Working Paper 23909. Series: Working Paper Series.

Fryer, Roland G. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *The Quarterly Journal of Economics*, 126(4): 1755–1798.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics*, 1(1): 112–135.

Hanushek, Eric A. 2020. "Education Production Functions." In *Economics of Education.* . 2 ed., , ed. Steve Bradley and Colin Green, 161–170. London:Academic Press.

Heckman, James. 1990. "Varieties of Selection Bias." *The American Economic Review*, 80(2): 313–318.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, 103(6): 2052–2086.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*, 2(3): 172–177.

Hirshleifer, Sarojini R. 2017. "Incentives for Effort or Outputs? A Field Experiment to Improve Student Performance." University of California at Riverside Working Paper.

Jacob, Brian A., Lars Lefgren, and David P. Sims. 2010. "The Persistence of Teacher-Induced Learning." *Journal of Human Resources*, 45(4): 915–943.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science*, 340(6130): 297–300.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics*, 91(3): 437–456.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies*, 76(3): 1071–1102.

Lee, Jong-Wha, and Hanol Lee. 2016. "Human capital in the long run." *Journal of Development Economics*, 122: 147–169.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies*, 70(2): 317–341.

Lipsey, Mark W., Kelly Puzio, Cathy Yun, Michael A. Hebert, Kasia Steinka-Fry, Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick. 2012. *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms.* National Center for Special Education Research.

Loerch, Andrew G. 2001. "Learning curves." In *Encyclopedia of Operations Research and Management Science.* , ed. Saul I. Gass and Carl M. Harris, 445–448. New York, NY:Springer US.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *The Quarterly Journal of Economics*.

Mbiti, Isaac, Mauricio Romero, and Youdi Schipper. 2019. "Designing Effective Teacher Perfor-

mance Pay Programs: Experimental Evidence from Tanzania." National Bureau of Economic Research Working Paper 25903.

NECTA. 2019. "Certificate of Secondary Education Examination (CSEE) Results (2010-2019)." The National Examination Council of Tanzania NECTA Open Data, Dar es Salaam, Tanzania.

PO-RALG. 2016. "Pre-Primary, Primary and Secondary Education Statistics in Brief." President's Office - Regional Administration and Local Government National Data, Dodoma, Tanzania.

Prendergast, Canice. 2011. "What Have We Learnt About Pay for Performance?" *The Economic and Social Review*, 42(No. 2): 113–134.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics*, 125(1): 175–214.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons. Google-Books-ID: bQBtw6rx_mUC.

Seo, Hee Kwon. 2016. "Pricing the Production of Mathematics Skill in Secondary Schools: Experimental Evidence from Tanzania." AEA RCT Registry.

Seo, Hee Kwon. 2017. "Do School Electrification and Provision of Digital Media Deliver Educational Benefits? First-year Evidence from 164 Tanzanian Secondary Schools."

Todd, Petra, and Kenneth I. Wolpin. 2018. "Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom." *Journal of Political Economy*, 126(6): 2608–2650.

World Bank. 2018. "World Development Report 2018: Learning to Realize Education's Promise." The World Bank, Washington, DC.

Tables

Table 1: Balance of Student Characteristics

	(1)	(2)	(3)	(4)	(5)
Statistic:	Control Mean	G1 - C	G2 - C	G3 - C	N
(*Statistic:*)	*(Control Sd.)*	*(Se.)*	*(Se.)*	*(Se.)*	*(%missing)*
A. Basic Demographics and Household Background, Feb. '16					
Female indicator	0.555	0.008	-0.023	0.019	6201
	(0.497)	(0.026)	(0.028)	(0.023)	(0)
Age	15.033	-0.142	-0.169**	-0.253***	5945
	(1.268)	(0.087)	(0.078)	(0.075)	(.0413)
Commute Distance (km)	3.189	-0.888***	-0.418	-0.387	5726
	(3.929)	(0.339)	(0.380)	(0.529)	(.0766)
Boarding student indicator	0.064	-0.053**	0.047	-0.001	6165
	(0.245)	(0.026)	(0.051)	(0.033)	(.00581)
Household has grid power	0.237	-0.019	0.000	0.012	6198
	(0.426)	(0.039)	(0.040)	(0.040)	(.000484)
Primary guardian finished secondary school or above	0.229	0.010	0.018	0.014	6159
	(0.420)	(0.025)	(0.024)	(0.023)	(.00677)
Primary guardian's occupation: Farming or Fishing	0.703	-0.042	-0.012	-0.033	5992
	(0.457)	(0.039)	(0.034)	(0.040)	(.0337)
Primary guardian's occupation: Technical or Managerial	0.097	-0.007	-0.024	-0.014	5992
	(0.296)	(0.017)	(0.015)	(0.015)	(.0337)
B. Educational Outlook, Preferences and Investments, Feb. '16					
Future occupation aimed for: Farming or Fishing	0.011	-0.004	-0.003	-0.004	5881
	(0.102)	(0.004)	(0.004)	(0.004)	(.0516)
Future occupation aimed for: Technical or Managerial	0.933	0.004	0.011	0.002	5881
	(0.251)	(0.014)	(0.013)	(0.013)	(.0516)
Intended area of focus: Science (not Arts or Commerce)	0.752	0.010	0.035	0.028	6127
	(0.432)	(0.027)	(0.027)	(0.027)	(.0119)
Likes math: Always (4 on a scale of 4)	0.604	0.018	-0.016	0.085*	6147
	(0.489)	(0.049)	(0.054)	(0.051)	(.00871)
Average Textbook Ownership (Nonmath Subjects)	0.071	-0.001	-0.015	-0.025	6198
	(0.141)	(0.017)	(0.016)	(0.015)	(.000484)
Textbook Ownership (Mathematics)	0.118	-0.008	-0.013	-0.038	6198
	(0.322)	(0.033)	(0.033)	(0.027)	(.000484)
Hrs/wk of non-math study after regular class hours	5.610	-0.194	-0.148	0.056	6139
	(2.630)	(0.262)	(0.240)	(0.252)	(.01)
Hrs/wk of math study after regular class hours	3.981	-0.079	-0.032	0.106	6132
	(2.586)	(0.220)	(0.220)	(0.230)	(.0111)
C. Mathematics Examination Results, Feb. '16					
Normalized mathematics marks, Feb. '16	-0.064	0.083	0.066	0.082	6197
	(0.924)	(0.079)	(0.103)	(0.089)	(.000645)
Pass rate (got 29.5 marks or above), Feb. '16	0.061	0.012	0.019	0.020	6197
	(0.240)	(0.015)	(0.020)	(0.016)	(.000645)
Pr. > Joint χ^2, All Treat. = 0			0.1802		

Note: Each row of coefficients is from a regression of the row variable on three treatment indicators (G1-G3). Last row: significance level of the joint test of orthogonality using a multinomial logit. Regressions include randomization-block (five-region) indicators. In parentheses: school-clustered standard errors. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10.

Table 2: Attendance on Dates of Examinations and O Level (Aggregate) Pass Indicator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Timing of Observation:	Year 1 (Oct. '16)		Year 2 (Oct. '17)		Year 3 (Oct. '18)		Year 3 (Nov. '18)
Explained Variable:	Attended	Enrolled	Attended	Enrolled	Attended	Enrolled	Graduated
Control (C) Mean Proportion	0.804	0.982	0.668	0.762	0.655	0.740	0.598
(Std. Dev.)	(0.397)	(0.135)	(0.471)	(0.426)	(0.476)	(0.439)	(0.490)
A. Treatment Variables							
Incentives Only (G1)	0.0388*	0.00977	0.0875***	0.0645**	0.0428	0.0605**	0.00674
	(0.0213)	(0.0131)	(0.0265)	(0.0256)	(0.0301)	(0.0282)	(0.0376)
	[0.0350]	[1]	[0.00400]	[0.0400]	[0.187]	[0.0670]	[1]
Technology Only (G2)	0.0500**	0.00279	0.0442*	0.0267	0.0315	0.0408	-0.0197
	(0.0213)	(0.0131)	(0.0247)	(0.0257)	(0.0293)	(0.0270)	(0.0359)
	[0.0350]	[1]	[0.0260]	[0.158]	[0.235]	[0.0670]	[1]
Both (G3)	0.0506**	0.00886	0.0603**	0.0395*	0.0572**	0.0490**	-0.00461
	(0.0219)	(0.0134)	(0.0241)	(0.0232)	(0.0249)	(0.0239)	(0.0314)
	[0.0350]	[1]	[0.0140]	[0.100]	[0.0740]	[0.0670]	[1]
B. Control Variables							
Age	-0.0199***	0.000218	-0.0278***	-0.0241***	-0.0311***	-0.0282***	-0.0549***
	(0.00471)	(0.00109)	(0.00535)	(0.00511)	(0.00599)	(0.00549)	(0.00671)
Year 0 (Feb. '16) Z-score	0.0462***	0.000621	0.0869***	0.0758***	0.0987***	0.0793***	0.171***
	(0.00603)	(0.00199)	(0.00833)	(0.00726)	(0.00826)	(0.00731)	(0.0104)
Commute Distance (km)	-0.00330	-0.00109	-0.00750***	-0.00751***	-0.00806***	-0.00732***	-0.00850***
	(0.00225)	(0.000799)	(0.00186)	(0.00201)	(0.00239)	(0.00203)	(0.00254)
Block (Five-region) FE	X	X	X	X	X	X	X
Observations	6,201	6,201	6,201	6,201	6,201	6,201	5,965
R-squared (Mean)	0.0390	0.00530	0.0869	0.0790	0.0878	0.0748	0.158
Clusters	170	170	170	170	170	170	170
F, Commute Distance = 0	2.161	-	16.32	-	11.40	-	-
Pr. > Joint F, All Treat. = 0	0.0925	0.483	0.0111	0.0805	0.146	0.154	0.905

Note: Difference-in-means coefficients. "Enrolled": enrolled in project school on exam date. "Graduated": obtained junior-secondary certificate at the end of the project period. Column (7) drops transferred students. Controls were missing at random for about 10% of students; estimates were obtained using multiple impuation and combined using Rubin's (1987) formulas. In parentheses: school-cluster-robust standard errors. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. In brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses).

Table 3: Effects on Performance

Timing of Observation:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Year 1 Z-score (Oct. '16)			Year 2 Z-score (Oct. '17)			Year 3 Z-score (Oct. '18)		
Specification:	Non-missing	Year 0 Controls	Selection Corrected	Non-missing	Year 0 Controls	Selection Corrected	Non-missing	Year 0 Controls	Selection Corrected
A. Treatment Variables									
Incentives Only (G1)	0.0732	0.0407	0.101	0.147	0.118	0.0541	0.120	0.105	0.101
	(0.0935)	(0.0716)	(0.0918)	(0.0944)	(0.0741)	(0.0935)	(0.0925)	(0.0737)	(0.0750)
	[0.628]	[0.393]	[0.321]	[0.138]	[0.0820]	[0.602]	[0.247]	[0.186]	[0.218]
Technology Only (G2)	0.134	0.110	0.175	0.137	0.122	0.0908	0.0589	0.0403	0.0295
	(0.129)	(0.0835)	(0.126)	(0.127)	(0.0745)	(0.0914)	(0.130)	(0.0805)	(0.0881)
	[0.628]	[0.393]	[0.321]	[0.222]	[0.0820]	[0.475]	[0.422]	[0.307]	[0.366]
Both (G3)	0.157	0.123	0.183*	0.415***	0.372***	0.331***	0.352***	0.301***	0.280***
	(0.103)	(0.0748)	(0.104)	(0.104)	(0.0767)	(0.0843)	(0.102)	(0.0781)	(0.0811)
	[0.628]	[0.393]	[0.321]	[0.00100]	[0.00100]	[0.00100]	[0.00300]	[0.00100]	[0.00300]
B. Linear Combinations of Estimators, Other Tests and Details									
$\beta_{G3} - \beta_{G1}$	0.0834	0.0819	0.0820	0.268**	0.253***	0.277***	0.232**	0.196**	0.179**
	(0.100)	(0.0668)	(0.0679)	(0.104)	(0.0876)	(0.0905)	(0.106)	(0.0870)	(0.0870)
$\beta_{G3} - \beta_{G2}$	0.0228	0.0122	0.00772	0.278**	0.250***	0.240***	0.293**	0.260***	0.251***
	(0.133)	(0.0793)	(0.0804)	(0.136)	(0.0897)	(0.0875)	(0.141)	(0.0942)	(0.0888)
$\beta_{G3} - \beta_{G2} - \beta_{G1}$	-0.0504	-0.0285	-0.0933	0.131	0.132	0.186	0.173	0.155	0.149
	(0.163)	(0.106)	(0.133)	(0.165)	(0.116)	(0.139)	(0.168)	(0.120)	(0.121)
Block (Five-region) FE	X	X	X	X	X	X	X	X	X
Observations	5,251	5,251	5,251	4,518	4,518	4,518	4,354	4,354	4,354
R-squared (Mean)	0.0536	0.570	0.574	0.0688	0.498	0.499	0.0566	0.473	0.478
Clusters	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.458	0.323	0.353	0.00150	0.000100	0.000700	0.00780	0.00210	0.00350

Note: Difference-in-means coefficients. Columns (2), (5) and (8) controls: year-0 score, age and commute distance. Column (3), (6) and (9) use Heckman's (1990) nonparametric control-function approach, using a (3rd-degree) polynomial of probit attrition-propensity score instrumented with commute distance. Controls were missing at random for about 10% of students; estimates were obtained using multiple imputation and combined using Rubin's (1987) formulas. Standard errors: clustered by school. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. Third row in brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses) as described in Anderson (2008).

Table 4: Effects on Effort (Reported Hours Per Week of Mathematics Study)

Timing of Observation:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Year 1 Math Study (Hrs/wk)			Year 2 Math Study (Hrs/wk)			Year 3 Math Study (Hrs/wk)		
Specification:	Non-missing	Year 0 Controls	Selection Corrected	Non-missing	Year 0 Controls	Selection Corrected	Non-missing	Year 0 Controls	Selection Corrected
Control (C) Mean	4.099	4.099	4.099	5.747	5.747	5.747	6.118	6.118	6.118
(Std. Dev.)	(2.979)	(2.979)	(2.979)	(5.703)	(5.703)	(5.703)	(6.193)	(6.193)	(6.193)
A. Treatment Variables									
Incentives Only (G1)	0.378	0.347	0.368	0.490	0.417	-0.277	0.602	0.616	0.587
	(0.345)	(0.349)	(0.377)	(0.480)	(0.454)	(0.560)	(0.545)	(0.517)	(0.514)
	[0.226]	[0.272]	[0.283]	[0.333]	[0.317]	[1]	[0.372]	[0.308]	[0.342]
Technology Only (G2)	0.522	0.500	0.517	0.490	0.462	0.114	0.289	0.289	0.279
	(0.402)	(0.400)	(0.444)	(0.551)	(0.492)	(0.529)	(0.592)	(0.550)	(0.543)
	[0.226]	[0.272]	[0.283]	[0.333]	[0.317]	[1]	[0.685]	[0.545]	[0.619]
Both (G3)	0.918**	0.887**	0.902**	1.662***	1.604***	1.153**	2.333***	2.296***	2.279***
	(0.383)	(0.381)	(0.414)	(0.558)	(0.516)	(0.563)	(0.615)	(0.594)	(0.604)
	[0.0570]	[0.0690]	[0.102]	[0.0110]	[0.00700]	[0.145]	[0.00100]	[0.00100]	[0.00100]
C. Linear Combinations of Estimators, Other Tests and Details									
$\beta_{G3} - \beta_{G1}$	0.540	0.540	0.534	1.172**	1.188**	1.431***	1.730***	1.680***	1.692***
	(0.374)	(0.371)	(0.368)	(0.512)	(0.488)	(0.497)	(0.585)	(0.573)	(0.579)
$\beta_{G3} - \beta_{G2}$	0.396	0.387	0.385	1.171**	1.142**	1.040*	2.043***	2.008***	2.001***
	(0.424)	(0.416)	(0.416)	(0.579)	(0.526)	(0.528)	(0.628)	(0.597)	(0.603)
$\beta_{G3} - \beta_{G2} - \beta_{G1}$	0.0181	0.0399	0.0170	0.681	0.725	1.317*	1.441*	1.392*	1.413*
	(0.547)	(0.543)	(0.568)	(0.754)	(0.699)	(0.756)	(0.829)	(0.789)	(0.786)
Observations	5,251	5,251	5,251	4,518	4,518	4,518	4,354	4,354	4,354
R-squared (Mean)	0.0884	0.102	0.103	0.0264	0.0616	0.0625	0.0549	0.0939	0.0951
Clusters	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.123	0.141	0.185	0.0276	0.0185	0.0322	0.00120	0.00100	0.00150

Note: Difference-in-means coefficients. Columns (2), (5) and (8) controls: year-0 score, age and commute distance. Column (3), (6) and (9) use Heckman's (1990) nonparametric control-function approach, using a (3rd-degree) polynomial of probit attrition-propensity score instrumented with commute distance. Controls were missing at random for about 10% of students; estimates were obtained using multiple impuation and combined using Rubin's (1987) formulas. Standard errors: clustered by school. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. Third row in brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses) as described in Anderson (2008).

Table 5: Year 2 and Year 3 Outcomes by Year 0 Performance Quintiles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Timing of Observation:	Year 2 Z-score (Oct. '17)					Year 3 Z-score (Oct. '18)				
Pretest Quintile:	Bottom	2nd	3rd	4th	Top	Bottom	2nd	3rd	4th	Top
Control (C) Mean	-0.731	-0.585	-0.444	-0.119	0.638	-0.588	-0.574	-0.484	-0.0816	0.712
(Std. Dev.)	(0.319)	(0.391)	(0.557)	(0.673)	(0.935)	(0.502)	(0.436)	(0.593)	(0.761)	(1.049)
A. Treatment Variables										
Incentives Only (G1)	0.205**	0.0975	0.0989	0.0271	-0.0322	0.107	0.115	0.233**	-0.00402	0.0375
	(0.0883)	(0.0885)	(0.0969)	(0.123)	(0.156)	(0.0768)	(0.0699)	(0.0987)	(0.118)	(0.149)
	[0.0230]	[0.374]	[0.448]	[1]	[0.387]	[0.199]	[0.114]	[0.0200]	[0.548]	[1]
Technology Only (G2)	0.0617	0.0520	0.00311	-0.00523	0.346*	0.0183	0.0462	0.0378	-0.136	0.119
	(0.0788)	(0.0771)	(0.0720)	(0.110)	(0.193)	(0.0830)	(0.0737)	(0.0723)	(0.114)	(0.201)
	[0.170]	[0.503]	[0.864]	[1]	[0.0810]	[0.380]	[0.216]	[0.251]	[0.309]	[1]
Both (G3)	0.250***	0.236***	0.375***	0.351***	0.449***	0.171**	0.267***	0.393***	0.227**	0.269*
	(0.0846)	(0.0792)	(0.0882)	(0.113)	(0.147)	(0.0834)	(0.0757)	(0.0942)	(0.111)	(0.151)
	[0.0110]	[0.0100]	[0.00100]	[0.00700]	[0.00800]	[0.144]	[0.00200]	[0.00100]	[0.147]	[0.300]
B. Linear Combinations of Estimators, Other Tests and Details										
β_{G3} - β_{G1}	0.0453	0.139*	0.276***	0.324***	0.481***	0.0640	0.152*	0.160	0.231**	0.231*
	(0.0867)	(0.0814)	(0.0983)	(0.105)	(0.167)	(0.0866)	(0.0907)	(0.117)	(0.108)	(0.139)
β_{G3} - β_{G2}	0.188**	0.184**	0.372***	0.356***	0.103	0.153*	0.221**	0.355***	0.363***	0.150
	(0.0832)	(0.0813)	(0.0840)	(0.0997)	(0.203)	(0.0882)	(0.0916)	(0.0914)	(0.102)	(0.192)
β_{G3} - β_{G2} - β_{G1}	-0.0164	0.0868	0.273**	0.329**	0.135	0.0457	0.106	0.122	0.367**	0.112
	(0.125)	(0.119)	(0.130)	(0.158)	(0.269)	(0.119)	(0.117)	(0.138)	(0.158)	(0.250)
Observations	4,518	4,518	4,518	4,518	4,518	4,354	4,354	4,354	4,354	4,354
R-squared (Mean)	0.505	0.505	0.505	0.505	0.505	0.481	0.481	0.481	0.481	0.481
Clusters	170	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.0102	0.0251	0.000100	0.00110	0.00380	0.138	0.00470	0.000100	0.00540	0.259

Note: Difference-in-means coefficients. Controls as in columns (6) and (9) of Table 3: age, year-0 score, randomization-block (five-region) indicators, and a (3rd-degree) polynomial of attrition-propensity score. Controls were missing at random for about 10% of students; estimates were obtained using multiple imputation and combined using Rubin's (1987) formulas. Standard errors: clustered by school in parentheses. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. In brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses).

Table 6: Other Potentially Relevant Inputs (Students' Technology Usage and Teachers' Time Input)

Timing of Observation:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year 2 Proxies for Inputs Usage (hrs/wk)				Year 3 Proxies for Inputs Usage (hrs/wk)			
Explained Variable:	Studied in School after 6pm	Printed Material Usage	ICT and Multimedia Usage	Total Teacher Hours	Studied in School after 6pm	Printed Material Usage	ICT and Multimedia Usage	Total Teacher Hours
Control (C) Mean	3.171	5.611	0.0381	6.404	8.465	6.372	0.135	6.681
(Std. Dev.)	(6.942)	(6.313)	(0.752)	(3.046)	(11.01)	(6.820)	(0.798)	(2.676)
A. Treatment Variables								
Incentives Only (G1)	-1.407	0.231	-0.0187	-0.158	-1.421	1.127*	-0.0837	-0.0650
	(1.328)	(0.702)	(0.0710)	(0.830)	(1.699)	(0.669)	(0.117)	(0.653)
	[1]	[0.329]	[0.360]	[1]	[1]	[0.0330]	[0.189]	[1]
Technology Only (G2)	0.931	2.494***	0.344***	0.292	1.229	2.032***	1.453***	0.394
	(1.516)	(0.611)	(0.111)	(0.714)	(1.762)	(0.643)	(0.266)	(0.579)
	[1]	[0.00100]	[0.00700]	[1]	[1]	[0.00200]	[0.00100]	[1]
Both (G3)	0.914	3.582***	0.235***	-0.0246	0.865	4.294***	1.321***	0.00369
	(1.176)	(0.643)	(0.0850)	(0.625)	(1.782)	(0.857)	(0.252)	(0.578)
	[1]	[0.00100]	[0.00700]	[1]	[1]	[0.00100]	[0.00100]	[1]
B. Linear Combinations of Estimators, Other Tests and Details								
β_{G3} - β_{G1}	2.321**	3.351***	0.254***	0.134	2.286	3.167***	1.404***	0.0687
	(0.924)	(0.691)	(0.0741)	(0.717)	(1.698)	(0.761)	(0.227)	(0.618)
β_{G3} - β_{G2}	-0.0172	1.088*	-0.109	-0.316	-0.365	2.262***	-0.132	-0.390
	(1.161)	(0.640)	(0.117)	(0.618)	(1.586)	(0.745)	(0.331)	(0.542)
β_{G3} - β_{G2} - β_{G1}	1.390	0.857	-0.0903	-0.158	1.056	1.135	-0.0483	-0.325
	(1.987)	(0.971)	(0.139)	(1.035)	(2.355)	(0.979)	(0.348)	(0.830)
Observations	4,518	4,518	4,518	170	4,354	4,354	4,354	170
R-squared (Mean)	0.0857	0.0827	0.0324	0.0951	0.152	0.0992	0.209	0.0547
Clusters	170	170	170	-	170	170	170	-
Pr. > Joint F, All Treat. = 0	0.0221	0	0	0.941	0.406	0	0	0.854

Note: Difference-in-means coefficients. Controls: age, year-0 score, randomization-block (region) indicators, and a polynomial of attrition-propensity score (degree 3). Controls were missing at random for about 10% of students; estimates were obtained by multiple impuation and combined via Rubin's (1987) formulas. "Total Teacher Hours" sum teaching, preparing and tutoring hours; regressions are weighted by classroom size. In parentheses: school-cluster-robust standard errors. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. In brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses).

Table 7: Selected Model Parameters and Simulated Treatment Effect Comparisons with Data

A. Model Parameters and Estimates			*B. Treatment Effects on Knowledge (Gross vs. Value-Added)*						*C. % Gaining Knowledge*	
[Statistic]	(1)	(2)	*[Statistic]*	(Data)	(1)	(2)	(1')	(2')	(2)	(2')
[Param.]			*[All, G.]*							
α_0	1.08	1.08	β_{G1}	5.41	-1.34	1.98	0.168	3.58	6.05	6.86
α_1	0.140	0.129	β_{G2}	9.08	11.5	6.03	10.4	5.37	4.39	4.58
γ_0	0.163	0.00100	β_{G3}	33.1	14.5	19.4	10.6	15.5	22.3	20.8
γ_1	0.00108	0.00208	$\beta_{G3\text{-}G2\text{-}G1}$	18.6	4.39	11.4	0.0253	6.52	11.9	9.32
δ_{cons}	0.148	0.0200								
$\Delta\delta_{K0}$	1.21E-04	3.33E-04	*[All, VA.]*							
τ_{cons}	0.0457	0.0437	β_{G1}	-	-0.241	3.03	0.144	3.50	-	-
τ_{SMR}	0.00677	0.0151	β_{G2}	-	10.6	5.56	10.4	5.44	-	-
$\pi_\$$	4.79	14.1	β_{G3}	-	11.6	16.74	10.6	15.52	-	-
π_{scale}	0.900	1.00	$\beta_{G3\text{-}G2\text{-}G1}$	-	1.16	8.16	0.232	6.58	-	-
θ_{cons}	1350	396								
θ_{STEM}	54.4	-33.7	*[Quintiles, G.]*							
p	2.58	3.58	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q5]*	13.5	2.23	10.1	0.0164	6.92	7.69	4.75
$\underline{E}$	1.41	5.61	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q4]*	32.9	11.6	22.2	0.0242	10.2	21.1	15.5
ι	1E-8	3.11	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q3]*	27.3	4.22	11.5	0.0293	7.23	14.0	12.4
σ_S	60.1	60.9	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q2]*	8.68	2.73	7.02	0.0297	4.55	9.38	8.23
$\underline{K}_0$	-	554	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q1]*	-1.64	0.944	3.38	0.0281	2.75	5.32	5.20
$\Delta\underline{K}_{0,SMR}$	-	-155								
$\underline{\pi}$	-	7.89	*[Quintiles, VA.]*							
			$\beta_{G3\text{-}G2\text{-}G1}$ *[Q5]*	-	0.451	8.26	0.0144	6.89	-	-
[% Baseline]			$\beta_{G3\text{-}G2\text{-}G1}$ *[Q4]*	-	2.62	13.5	0.0220	10.3	-	-
Disinterested	-	86.1	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q3]*	-	1.16	8.33	0.0271	7.33	-	-
Lost	-	78.0	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q2]*	-	0.874	5.32	0.0275	4.64	-	-
Both D. & L.	-	77.0	$\beta_{G3\text{-}G2\text{-}G1}$ *[Q1]*	-	0.339	2.90	0.0265	2.81	-	-
Neither	100	12.8								

Note: Model (1) refers to the benchmark model; (2), the the productivity-thresholds model. Top of panel A reports selected parameter estimates. In the bottom of panel A, "disinterested" refers to students whose valuation of knowledge fell short of minimum interest threshold; "lost" refers to students whose entering grade-level knowledge fell short of minimum preparedness threshold; "both" refers to both disinterested and lost; "neither" refers to neither disinterested nor lost. In panel B, column (Data) reproduces estimates from column (6) of table 3 and the bottom row of table 5. Columns (1) and (2) report differences in simulated means across treatment groups, averaged across 1,000 simulation draws, leaving the original balance of student characteristics across treatment groups intact. Columns (1') and (2') report averaged differences in simulated means across treatments, assuming that the whole sample gets shocked with each treatment. In panel C, columns (2) and (2') report differences in model-implied proportions of students who are gaining knowledge—i.e., those who are neither disinterested nor lost.

Table 8: Revealed-Preference–Based Valuations of Achievement and Inputs across Models

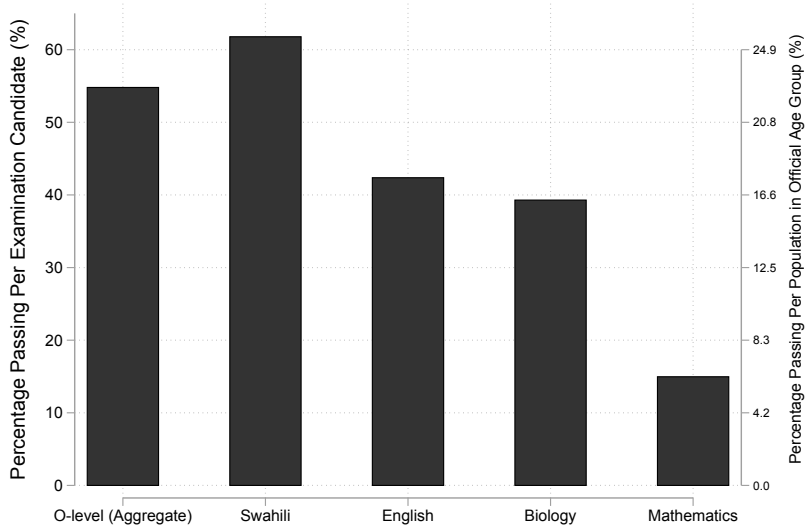
	A. Mean Valuation Estimates ($ per good)				*B. Mean Cost Estimates ($ per good)*		
[Sample]	(1)	(2)	(2; entry cost)	(2; 0 if $MB < \pi$)	(Data)	(1)	(2)
[All]							
+1 Ki (0.01σ)	1.32	0.42		0.0801	-	1.42	0.0230
+10 Ki (0.10σ)	13.2	4.20		0.801	-	71.7	3.26
+20 Ki (0.20σ)	26.4	8.41	(same as left)	1.60	-	17100	169
+10 Ki given G2	13.8	4.31		1.01	-	54.55	2.048
$E[\theta]$	289	26.6		-	-	-	-
$E[\theta\phi/\sigma] \times 10$	3.23	0.272		-	-	-	-
Incentives (G1)	5.61	8.13	2.46	-	5.30	5.06	4.99
Technology (G2)	31.0	5.24	1.01	-	6.13	(same as left)	
Both (G3)	37.2	17.8	5.11	-	13.1	11.8	11.9
Complementarity	0.598	4.44	1.63	-	1.72	0.605	0.792
[Q5]							
+10 Ki (0.10σ)	21.9	5.12		2.79	-	70.9	14.1
+20 Ki (0.20σ)	43.7	10.2	(same as left)	5.58	-	1680	721
+10 Ki given G2	23.1	5.25		3.19	-	61.9	8.42
$E[\theta\phi/\sigma] \times 10$	10.2	0.885		-	-	-	-
Incentives (G1)	12.0	22.1	9.5	-	10.5	11.4	11.9
Technology (G2)	98.4	17.2	1.69	-	6.13	(same as left)	
Both (G3)	111	46.3	11.9	-	19.9	18.4	19.3
Complementarity	1.03	7.06	0.71	-	3.22	0.880	1.31
[Q4]							
+10 Ki (0.10σ)	12.6	4.26		0.610	-	63.0	0.576
+20 Ki (0.20σ)	25.2	8.51	(same as left)	1.22	-	6350	41.2
+10 Ki given G2	13.5	4.43		0.956	-	50.7	0.762
$E[\theta\phi/\sigma] \times 10$	2.96	0.24		-	-	-	-
Incentives (G1)	5.76	7.23	1,22	-	5.90	5.49	5.22
Technology (G2)	17.1	4.62	1.76	-	6.13	(same as left)	
Both (G3)	23.5	17.9	5.92	-	13.4	12.3	12.5
Complementarity	0.643	6.06	2.94	-	1.35	0.665	1.15

Note: [From top to bottom:] Rows labeled "+s Ki" refer to adding s scaled-score units of knowledge given the status quo learning environment. "+10 Ki given G2" refers to +10 scaled-score units given additionally the experimental technology (G2). "E[θ]" refers to promotion; "$E[\theta\phi/\sigma] \times 10$" refers to how much the chance of promotion affects the valuation of adding 10 scaled-score units. The rest of the rows refer to valuations of the treatment interventions and their interaction effect. Model (1) refers to the benchmark model; (2), the model with productivity traps. Column labeled "(2; entry cost)" shows average valuations interpreting minimum interest thresholds as stemming from entry costs of learning new material on the curriculum. Column labeled "(2; 0 if MB < π)" shows "perceived" estimates as if students whose interests are lower than the minimum interest threshold valued each additional scaled-score unit at 0. In panel B, column (Data) shows mean cost of each treatment as seen in program data; columns (1) and (2) show simulated effort costs and program costs across models.

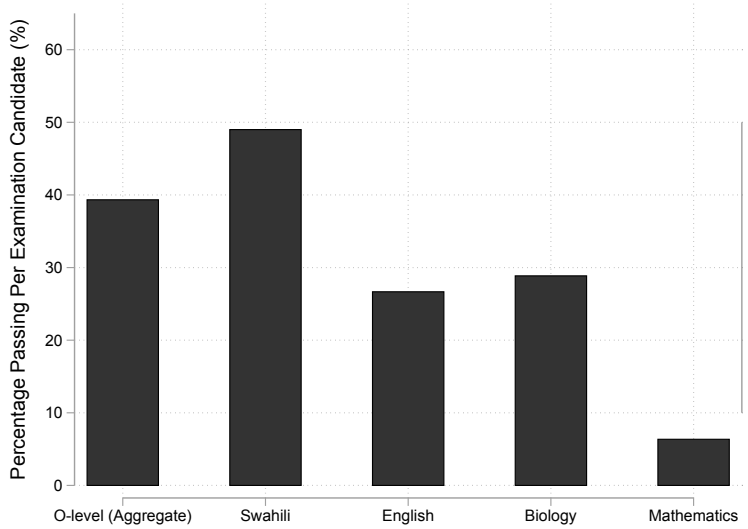
Figures

Figure 1: O Level Pass Rates for Secondary School Students (2012–2015)

(a) Tanzania



(b) Sample Schools

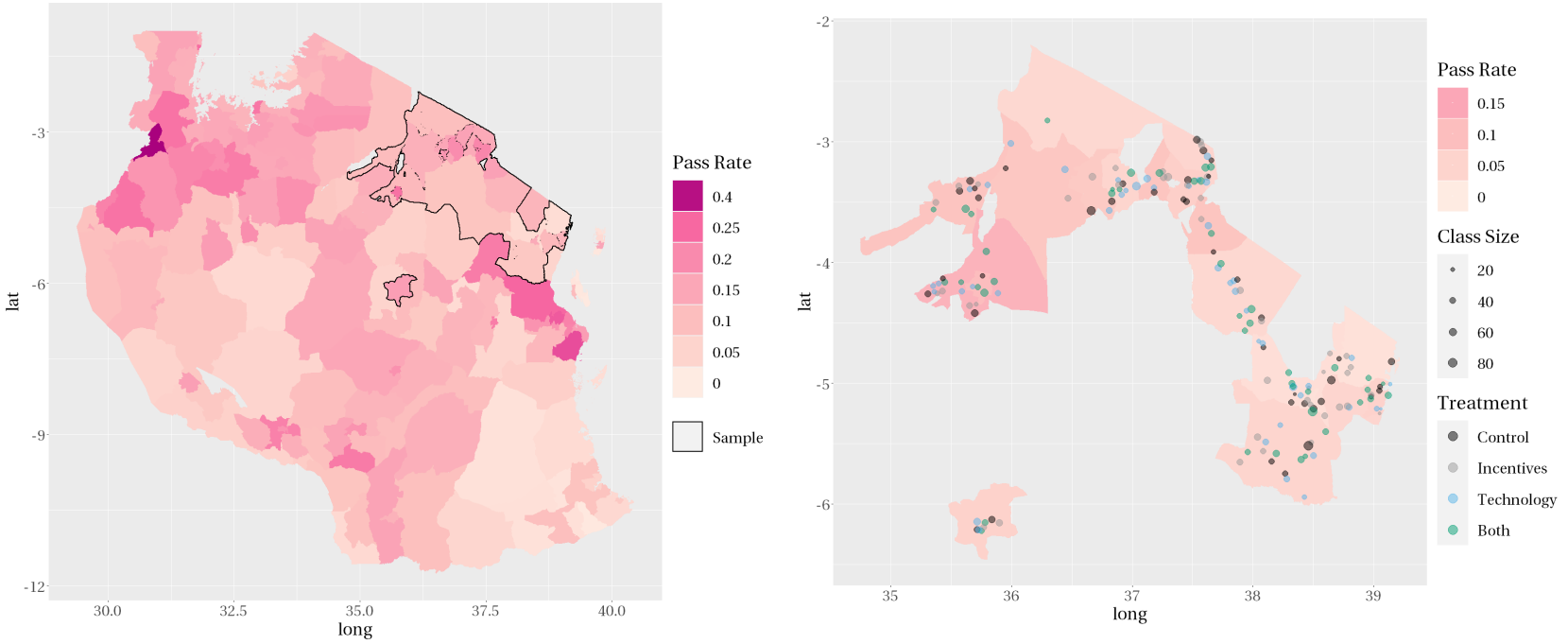


Note: Author's calculations using government data (PO-RALG, 2016; NECTA, 2019). Figure 1a plots the numbers of students who passed over those sat for the O Levels between 2012 and 2015; approximately 1.4 million students sat for these examinations over these four years. The right-hand-size y-axis plots the numbers of students who passed over the number of youths who belonged in the official secondary-school age group population (approximately 3.4 million youths). O Level certification requires obtaining at least two D's out of five required subjects and two (or more) optional subjects; the five required subjects include Swahili, English, Civics, Biology and Mathematics—the pass rate for Civics look similar to Swahili and English pass rates. Figure 1a plots the numbers of students who passed in the SMR sample schools over the number of students who sat for the examinations across these four years (44,804 students).

Figure 2: National Mathematics Performance and Characteristics of Project Schools (2015)

(a) National District Pass Rates

(b) Sample District Pass Rates



Note: Author's calculations using government and project survey data (NECTA, 2019). The research team initially targeted all schools without electricity in 23 northern Tanzanian districts. Districts shown in fig. 2b are three fewer than those demarcated in black in fig. 2a, because some districts were found with no unelectrified school and dropped.

Figure 3: Incentives, Textbooks, Solar Lights and Solar TVs

(a) Incentives

(b) Technology (books)

(e) Technology (videos)

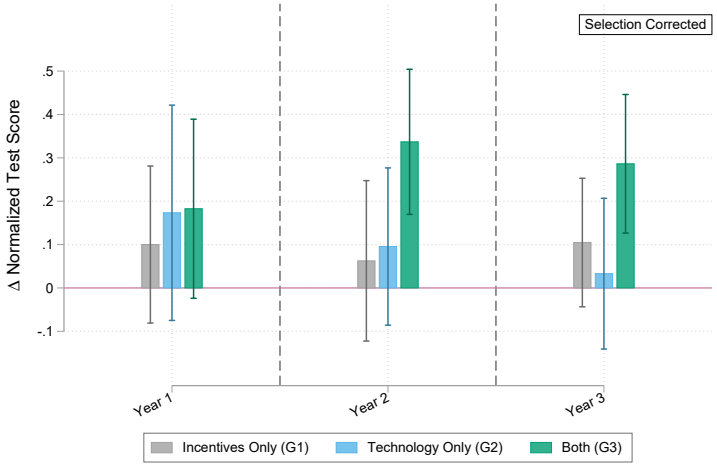
(c) Technology (lights)

(d) Technology (TV)

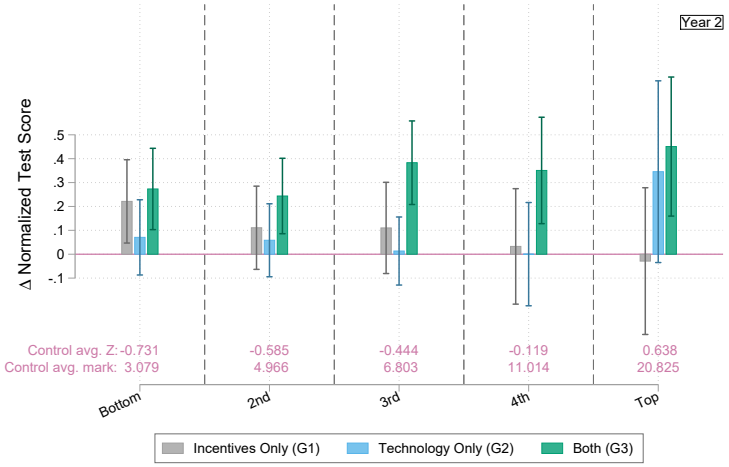
Note: Sub-figure (c): M120's came with 3 different types of lights, 6 lights per system, 12 lights in total across two systems provided. These systems covered approximately two large classrooms and one office. Sub-figure (d): The systems also came with one 16" and one 19" solar TV. Shown in sub-figure (e) is an example classroom viewing a video.

Figure 4: Effects on Test Performance

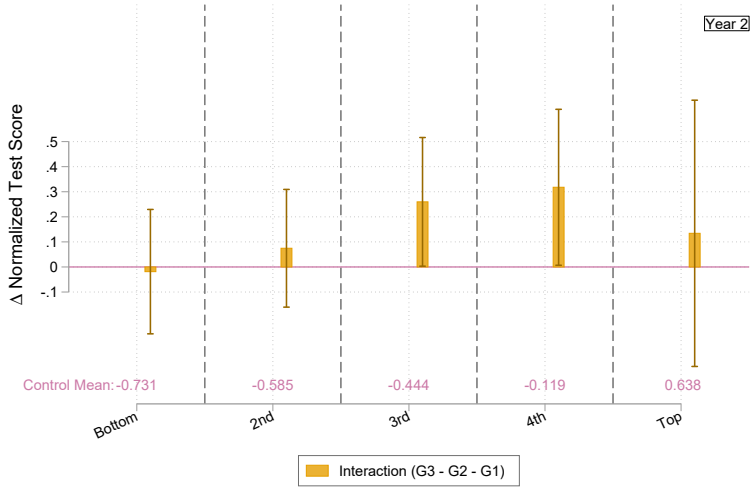
(a) Effects across Years



(b) Effects by Quintiles



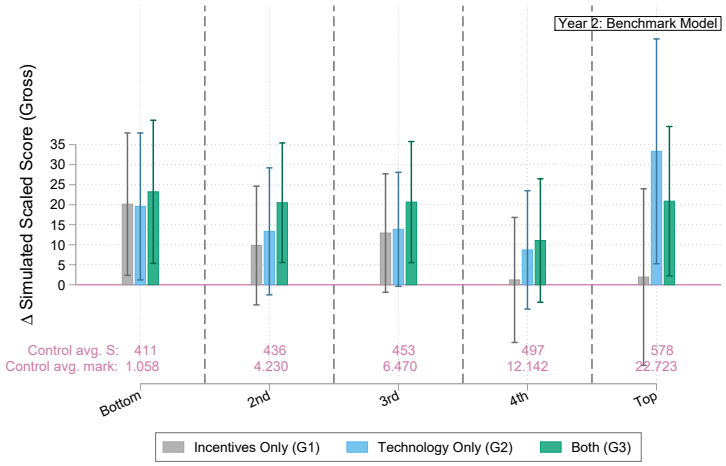
(c) Interaction by Quintiles



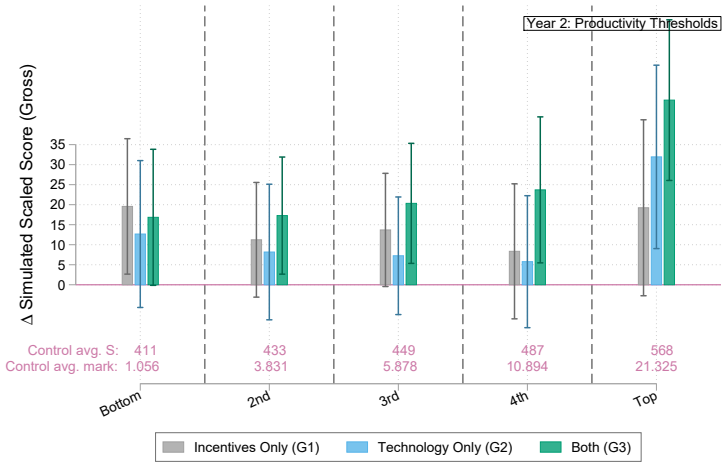
Note: Difference-in-means coefficients. Regressions control for year-0 score, age, randomization-block (five-region) indicators, and a (3rd-degree) polynomial of probit attrition-propensity score instrumented with commute distance (Heckman, 1990). Controls were missing at random for about 10% of students; estimates were obtained using multiple imputation and combined using Rubin's (1987) formulas. Results are robust to exclusion of controls and/or missing cases (c.f. table 3); heterogeneous results are similar in year 3 (c.f. table 5). School-cluster-robust 95%-confidence intervals indicated.

Figure 5: Model-Generated Effects by Pretest Performance Quintiles: Benchmark Model vs. Productivity-Thresholds Model

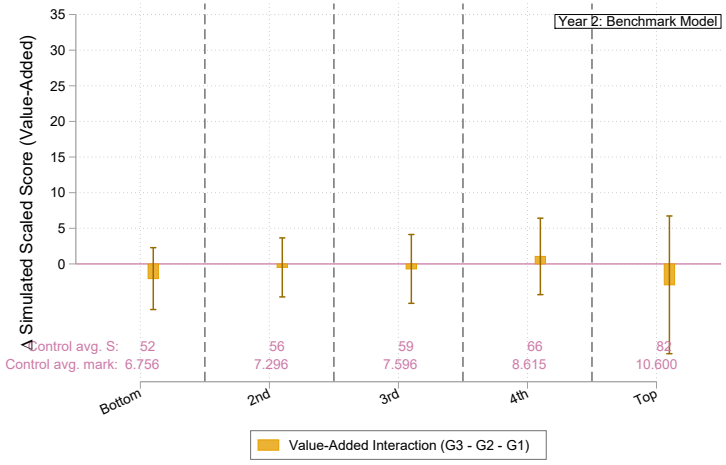
(a) Benchmark Model: Effects by Quintiles



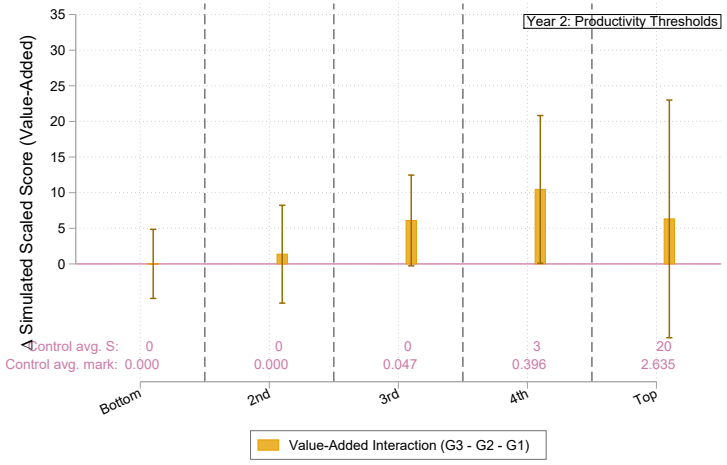
(c) Productivity-Thresholds Model: Effects by Quintiles



(b) Benchmark Model: Value-Added Interactions



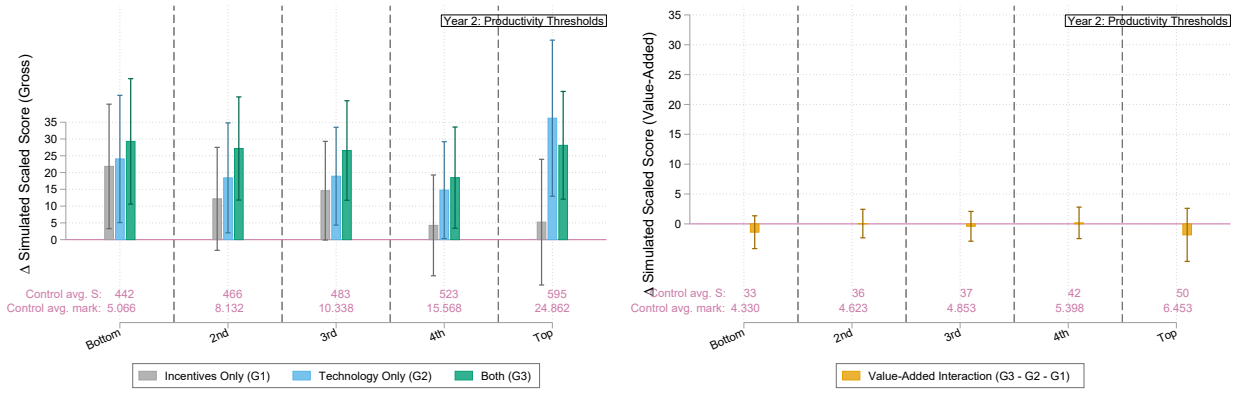
(d) Productivity-Thresholds Model: Value-Added Interactions



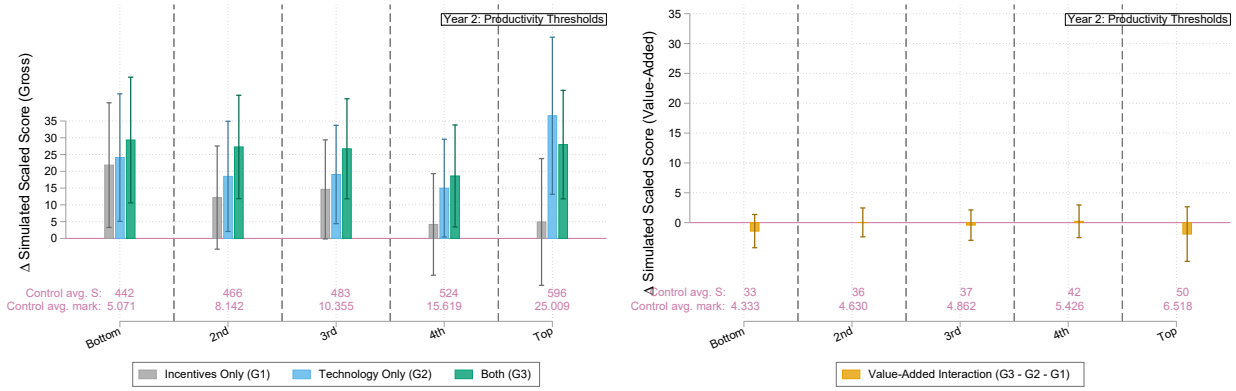
Note: Simulated based on structurally estimated models of knowledge accumulation, allowing for individual-level unobserved heterogeneity. In the "benchmark model," students balance knowledge returns against convex effort costs, given value-added multiplicative in exogenous inputs, effort, and entering grade-level knowledge ("preparedness"). In the "productivity-thresholds model," students additionally face threshold costs of learning new material involving minimum interest and preparedness levels. (C.f. reduced-form treatment effects in fig. 4 were NOT targeted during estimation, and as such can be viewed as yardsticks of model performance; 1 normalized-test-score unit equals 100 scaled-score units.) School-cluster-robust 95%-confidence intervals indicated.

Figure 6: Productivity-Thresholds Model Results by Different Combinations of Included Structural Components

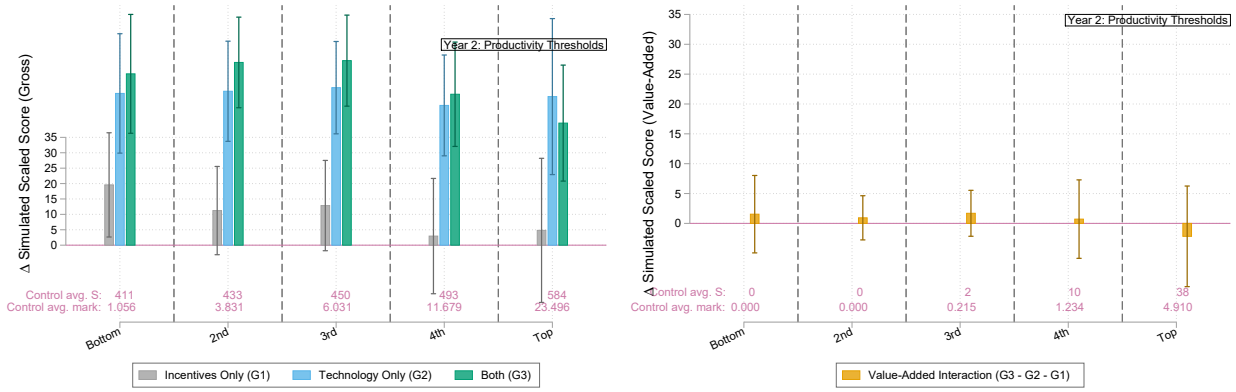
(a) Scenario 1 (S1): NO promotion concern & NO knowledge threshold & NO interest threshold



(b) Scenario 2: S1 + promotion concern

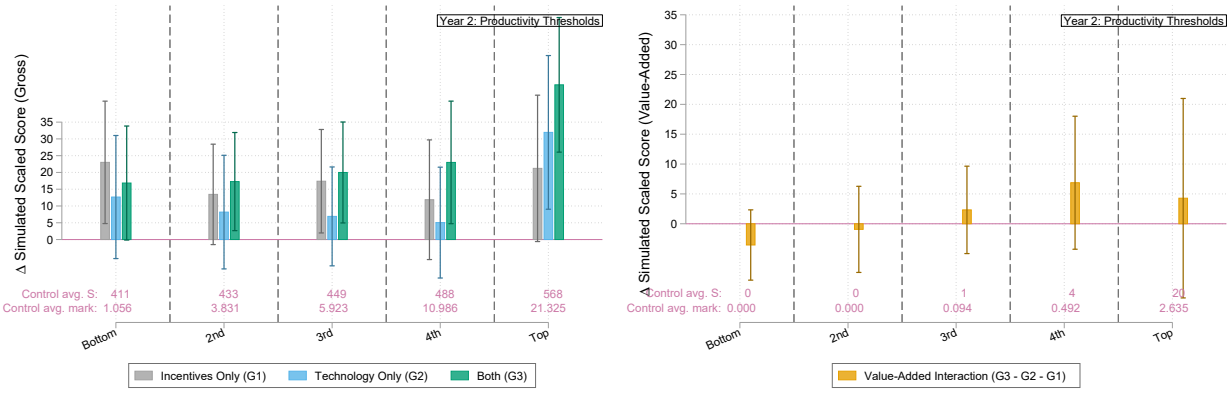


(c) Scenario 3: S1 + promotion concern + knowledge threshold

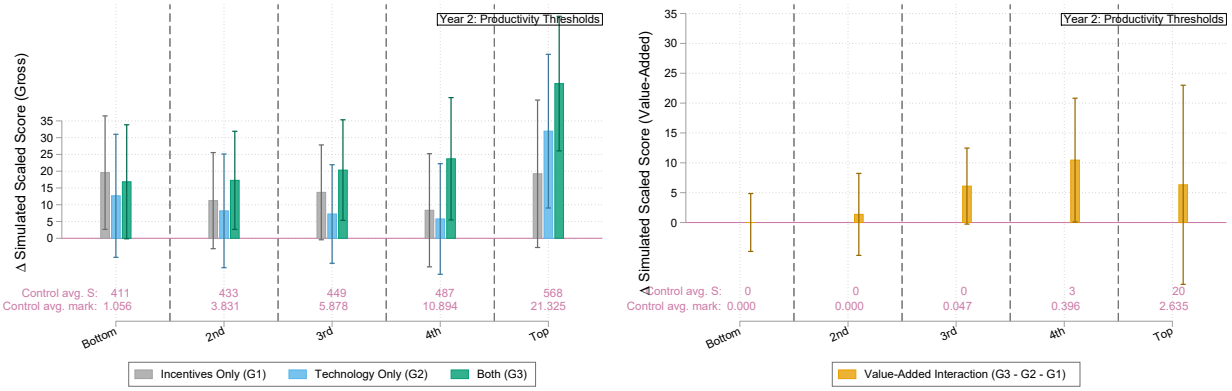


Note: (Subfigures continue on the next page. See the end of next page for figure notes.)

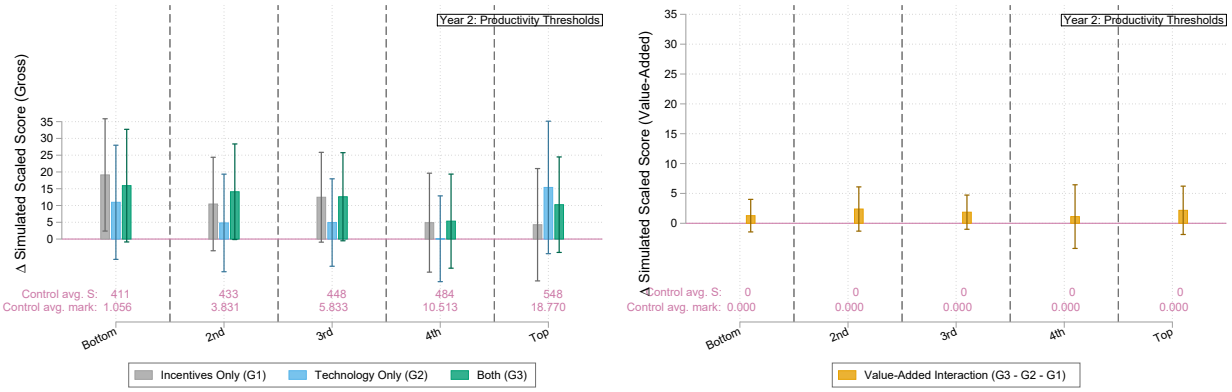
(d) Scenario 4: S1 + promotion concern + interest threshold



(e) Scenario 5: S1 + promotion concern + knowledge threshold + interest threshold

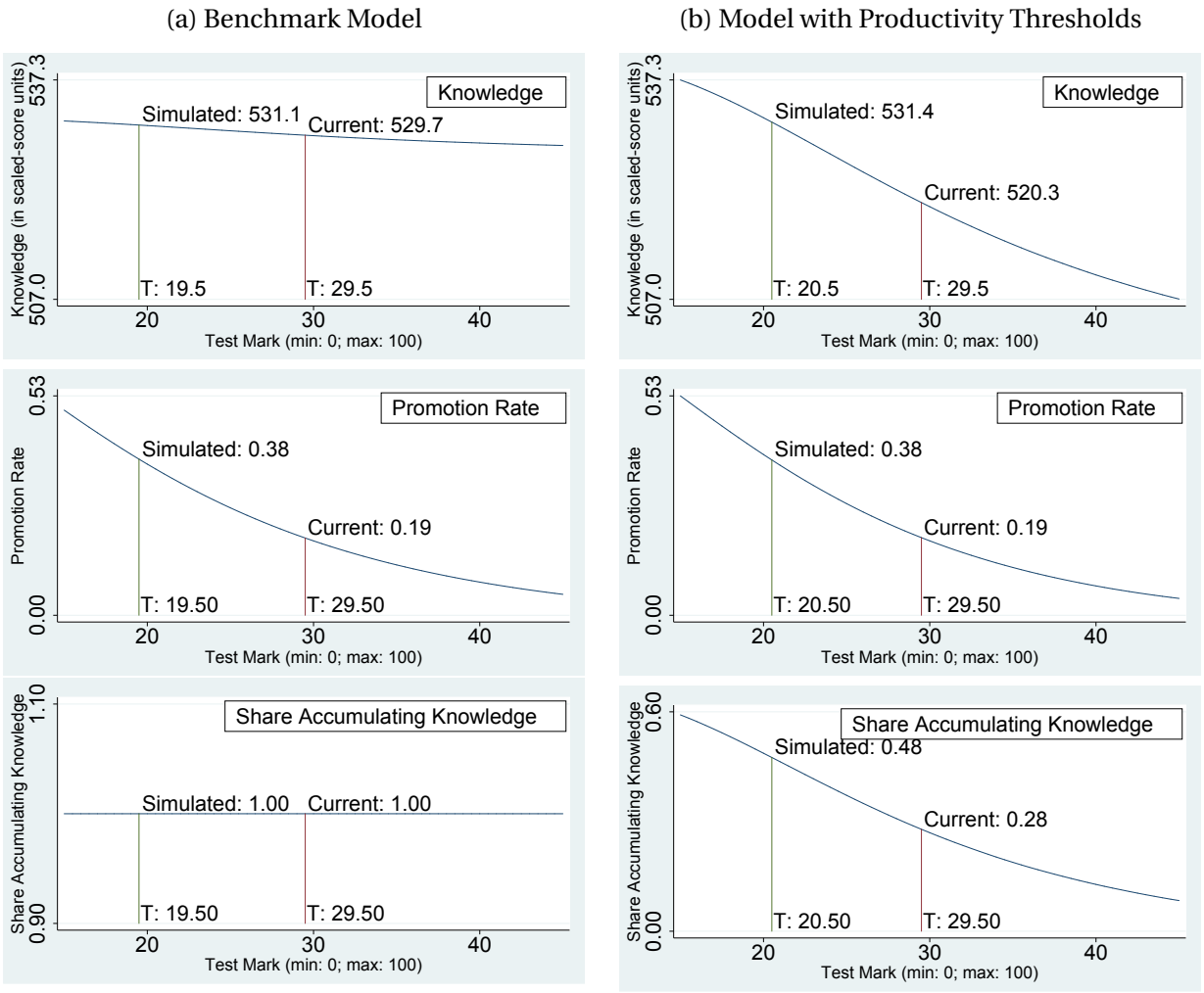


(f) Scenario 6: S1 + knowledge threshold + interest threshold



Note: (Continued from the previous page.) Difference-in-means coefficients, simulated based on structurally estimated models of student learning allowing for individual-level unobserved factor heterogeneity (see sections 3 and 4). In Scenario 1, "NO promotion concern" means $\theta = 0$; "NO knowledge threshold" means $\underline{K}_0 = 0$; "NO interest threshold" means $\underline{\pi} = 0$. Ensuing scenarios are based on adding different combinations of structural components as labeled. School-cluster-robust 95%-confidence intervals indicated.

Figure 7: Outcomes Given Counterfactual Promotion Cutoffs and Technology Provision



Note: "Promotion cutoffs" refer to minimum test marks a student must achieve as a prerequisite to training in STEM-related occupations. Plots indicate endogenous, equilibrium simulation outcomes that vary in counterfactual promotion cutoffs, conditional on distributing the experimental technology nationally. Indicated in red is the current (status-quo) cutoff; indicated in green is the cutoff that would double the number of students passing endogenously. Each sample observation is weighted nationally using the ratio between a Weibull density fitted to Form Two National Assessment grade distribution and the density of year-1 mock-test marks. "Share Accumulating Knowledge" refers to the proportion of students who meaningfully accumulate knowledge in the curriculum (i.e., those who are "paying the entry cost," or "not giving up" on learning the content).

A Appendix Tables

Table A1: A Review of Five Selected Student Performance Subsidy Experiments

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Study	Setting	Unit Period [All Periods]	Subject(s)	Grade(s)	Unit Subsidy†		Technology / Inputs	Average Treatment Effect(s)‡			
					$	% GDPPC		Incentives Only	Technology Only	Both	Incentives | Tech.
Fryer (2011)	New York	2 Months [1 Year]	Reading, Math	4, 7	$0.20~$0.40	>0.001%	-	−0.12σ ~ −0.05σ			
Fryer (2011)	Chicago	5 Weeks [1 Year]	Core Courses	9	$0.50	0.001%	-	0.09σ			
Behrman et al. (2015)	Mexico	1 Year [3 Years]	Math	10, 11, 12	$8	0.1%	Teacher Incentives	0.17σ ~ 0.32σ	−0.05σ ~ 0.14σ	0.19σ ~ 0.63σ	0.23σ ~ 0.53σ
Hirshleifer (2017)	Mumbai, Pune	40 Days [80 Days]	Math	4, 5, 6	$0.03	0.002%	Tablet, Softwares				0.24σ
This Study (2018)	Northern Tanzania	1 Year [3 Years]	Math	9, 10, 11	$0.50	0.05%	Solar Energy, Bilingual Textbooks, Videos	-0.04σ ~ 0.05σ	-0.06σ ~ 0.08σ	0.13σ ~ 0.28σ	0.05σ ~ 0.30σ

Note: Author's compilation. Selection was based on whether student incentives in the experiment could be approximated as "unit subsidies": piece-rate payment contracts per percentage mark on period-end achivement test or report card. In each experiment, students were randomized into groups receiving Incentives Only, Technology (or Inputs) Only, Both (Incentives and Technology), or none of the above. Treatments were delivered in the beginning of each period over multiple periods. Columns (9)-(11) show the range of reported period-cohort-specific difference-in-means treatment effects. Column (12) reports differences between the Both treatment effect and Technology Only treatment effect; hence, column (12) identifies the effectiveness of the incentive when technology were provided to students, whereas by comparison column (9) identifies the effectiveness of the incentive when the technology was not provided to students. Hirshleifer (2017) did not include a control group that did not receive any treatment beyond what is available in the normal schooling environment (but included Technology Only, Both, and an additional treatment group that received subsidies on input usage whose effect is beyond the scope of this paper and is ommitted); in her case, the Both against Technology Only treatment effect identifies the analogous difference for column (12).

† Columns (6) and (7) refer to approximate size of performance contract promised to students per end-of-period percentage mark. Values in column (7) are obtained by dividing those in column (6) by the respective national Gross Domestic Product per capita.

‡ In Behrman et al. (2015), "Both" treatment was not a simple sum of "Incentive" and "Inputs" treatments, but students additionally received rewards from peer student performance, teachers from peer-teacher performance, and school administrators from school-wide performance.

Table A2: Minimum Detectable Effect Size Calculations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Timing of Observation:	Year 0 (Feb. '16)		Year 1 (Oct. '16)		Year 2 (Oct. '17)		Year 3 (Oct. '18)	
Power:	80%	80%	80%	80%	80%	80%	80%	80%
MHT Correction:	None	Bonferroni	None	Bonferroni	None	Bonferroni	None	Bonferroni
Outcome Variables:								
Normalized Mathematics Marks	0.237	0.274	0.203	0.234	0.231	0.267	0.224	0.258
Parameters and Moments:								
Number of Hypotheses	1	3	1	3	1	3	1	3
J_{T} (Treatment Size)	43	43	43	43	43	43	43	43
J_{C} (Control Size)	40	40	40	40	40	40	40	40
n (Cluster Size)	36.453	36.453	30.888	30.888	26.576	26.576	25.612	25.612
τ_{α/2}	1.96	2.39	1.96	2.39	1.96	2.39	1.96	2.39
τ_{1-κ}	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
p (Treatment Share)	0.518	0.518	0.518	0.518	0.518	0.518	0.518	0.518
c (Compliance among Treated)	1	1	1	1	1	1	1	1
s (Defiance among Control)	0	0	0	0	0	0	0	0
ρ (Residual Intracluster Corr.)	0.138	0.138	0.232	0.232	0.250	0.250	0.222	0.222
σ (Residual Std. Dev.)	0.958	0.958	0.650	0.650	0.711	0.711	0.724	0.724

Note : This table reports minimum detectable effect sizes (MDEs) calculated under different clustered-randomized-design scenarios, using equation (12) of Duflo et al. (2007) (Bloom (2005)). The Sharpening Mathematics Review School Program randomized 170 9th-grade classrooms, each sampled from a distinct school, into three treatment groups and a control group, with approximately 43 schools in each treatment group and 40 in control, targeting normalized mathematics marks on follow-up curriculum-based tests. Odd columns report MDE with alpha unadjusted for multiple-hypotheses testing; evens report MDEs with alpha adjusted for three independent hypotheses using Bonferroni. The moments used include the average number of students (n), the intracluster correlation coefficients (ρ) and standard deviations (σ) of normalized mathematics test scores. In columns (1)-(2), moments are calculated based on year 0 (Feb. '16) F1 (grade 8) results; power reported controls for age, commute distance and randomization-block (five-region) indicators. Columns (3)-(4) report the power realized on year 1 (Oct. '16) F2 (grade 9) results, controlling for age, commute distance, randomization-block (five-region) indicators and year-0 marks (administered before the program's incentive contracts were announced and before year-1 math curriculum textbooks and videos were delivered). Columns (5)-(6) report the analogous power for year 2 (Oct. '17) F3 (grade 10) results; Columns (7)-(8), for year 3 (Oct. '18) F4 (grade 11) results.

Table A3: Effects on Performance (Observation Missing Controls Dropped)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing of Observation:	Year 1 Z-score (Oct. '16)			Year 2 Z-score (Oct. '17)			Year 3 Z-score (Oct. '18)		
Specification:	Non-missing	Feb. '16 Controls	Selection Corrected	Non-missing	Feb. '16 Controls	Selection Corrected	Non-missing	Feb. '16 Controls	Selection Corrected
A. Treatment Variables									
Incentives Only (G1)	0.0727	0.0413	0.0503	0.147	0.0789	-0.0433	0.119	0.0657	0.0401
	(0.0935)	(0.0756)	(0.0875)	(0.0944)	(0.0696)	(0.0804)	(0.0926)	(0.0721)	(0.0708)
	[0.619]	[0.641]	[1]	[0.138]	[0.244]	[1]	[0.249]	[0.572]	[0.617]
Technology Only (G2)	0.134	0.0859	0.0817	0.137	0.0691	0.0134	0.0587	-0.0248	-0.0562
	(0.129)	(0.0808)	(0.0985)	(0.127)	(0.0656)	(0.0703)	(0.130)	(0.0674)	(0.0677)
	[0.619]	[0.407]	[1]	[0.222]	[0.244]	[1]	[0.426]	[0.908]	[0.617]
Both (G3)	0.157	0.136*	0.127	0.415***	0.371***	0.283***	0.349***	0.309***	0.241***
	(0.102)	(0.0811)	(0.0984)	(0.104)	(0.0797)	(0.0816)	(0.102)	(0.0832)	(0.0824)
	[0.619]	[0.397]	[1]	[0.00100]	[0.00100]	[0.00300]	[0.00300]	[0.00100]	[0.0120]
B. Linear Combinations of Estimators, Other Tests and Details									
β_{T3} - β_{T1}	0.0842	0.0950	0.0770	0.268**	0.292***	0.326***	0.230**	0.244***	0.201**
	(0.100)	(0.0715)	(0.0711)	(0.104)	(0.0886)	(0.0901)	(0.106)	(0.0872)	(0.0873)
β_{T3} - β_{T2}	0.0232	0.0504	0.0457	0.278**	0.302***	0.269***	0.290**	0.334***	0.297***
	(0.133)	(0.0761)	(0.0759)	(0.136)	(0.0863)	(0.0849)	(0.141)	(0.0839)	(0.0832)
β_{T3} - β_{T1} - β_{T2}	-0.0496	0.00908	-0.00467	0.131	0.223**	0.312***	0.171	0.268**	0.257**
	(0.163)	(0.107)	(0.115)	(0.165)	(0.112)	(0.119)	(0.168)	(0.111)	(0.111)
Block (Five-region) FE	X	X	X	X	X	X	X	X	X
Observations	5,251	4,697	4,697	4,518	4,079	4,079	4,351	3,919	3,919
R-squared	0.054	0.547	0.551	0.069	0.483	0.485	0.057	0.461	0.466
Clusters	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.456	0.356	0.588	0.00140	0.000100	0.00140	0.00860	0.000700	0.00470

Note: Difference-in-means coefficients. Columns (2), (5) and (8) control for year-0 score, age and commute distance. Column (3), (6) and (9) use Heckman's (1990) nonparametric control-function approach, using a (degree-3) polynomial of probit attrition-propensity score instrumented with commute distance. Standard errors: clustered by school. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. Third row in brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses) as described in Anderson (2008).

Table A4: Reported Hours Per Week of Mathematics Study (Observation Missing Controls Dropped)

Timing of Observation:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Year 1 Math Study (hrs/wk)			Year 2 Math Study (hrs/wk)			Year 3 Math Study (hrs/wk)		
Specification:	Non-missing	Feb. '16 Controls	Selection Corrected	Non-missing	Feb. '16 Controls	Selection Corrected	Non-missing	Feb. '16 Controls	Selection Corrected
Control (C) Mean	4.099	4.149	4.149	5.747	5.783	5.783	6.118	6.080	6.080
(Std. Dev.)	(2.978)	(3.002)	(3.002)	(5.701)	(5.757)	(5.757)	(6.191)	(6.060)	(6.060)
A. Treatment Variables									
Incentives Only (G1)	0.382	0.401	0.322	0.490	0.398	-0.429	0.603	0.567	0.542
	(0.346)	(0.356)	(0.378)	(0.480)	(0.459)	(0.563)	(0.545)	(0.516)	(0.514)
	[0.220]	[0.211]	[0.395]	[0.333]	[0.602]	[0.809]	[0.371]	[0.378]	[0.414]
Technology Only (G2)	0.522	0.469	0.352	0.490	0.284	-0.103	0.289	0.156	0.136
	(0.402)	(0.414)	(0.440)	(0.551)	(0.490)	(0.512)	(0.592)	(0.535)	(0.529)
	[0.220]	[0.211]	[0.395]	[0.333]	[0.602]	[1]	[0.684]	[0.698]	[0.782]
Both (G3)	0.915**	1.033***	0.912**	1.662***	1.799***	1.193**	2.326***	2.352***	2.328***
	(0.383)	(0.388)	(0.407)	(0.558)	(0.522)	(0.573)	(0.615)	(0.582)	(0.604)
	[0.0580]	[0.0270]	[0.0870]	[0.0110]	[0.00300]	[0.133]	[0.00100]	[0.00100]	[0.00100]
C. Linear Combinations of Estimators, Other Tests and Details									
β_{T3} - β_{T1}	0.532	0.631*	0.590	1.172**	1.401***	1.622***	1.724***	1.786***	1.786***
	(0.374)	(0.376)	(0.369)	(0.512)	(0.497)	(0.501)	(0.587)	(0.584)	(0.600)
β_{T3} - β_{T2}	0.393	0.564	0.559	1.171**	1.515***	1.296**	2.038***	2.197***	2.192***
	(0.424)	(0.428)	(0.426)	(0.579)	(0.526)	(0.530)	(0.629)	(0.596)	(0.610)
β_{T3} - β_{T1} - β_{T2}	0.0108	0.163	0.238	0.681	1.117	1.725**	1.435*	1.630**	1.650**
	(0.547)	(0.558)	(0.574)	(0.754)	(0.703)	(0.737)	(0.829)	(0.788)	(0.790)
Observations	5,251	4,697	4,697	4,518	4,079	4,079	4,351	3,919	3,919
R-squared	0.088	0.102	0.102	0.026	0.060	0.061	0.055	0.101	0.102
Clusters	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.124	0.0719	0.156	0.0275	0.00470	0.0122	0.00130	0.000400	0.000900

Note : Difference-in-means coefficients. Columns (2), (5) and (8) controls: year-0 score, age and commute distance. Column (3), (6) and (9) use Heckman's (1990) nonparametric control-function approach, using a (degree-3) polynomial of probit attrition-propensity score instrumented with commute distance. Standard errors: clustered by school. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. Third row in brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses) as described in Anderson (2008).

Table A5: Year 2 and Year 3 Outcomes by Year 0 Performance Quintiles (Observation Missing Controls Dropped)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Timing of Observation:	Year 2 Z-score (Oct. '17)					Year 3 Z-score (Oct. '18)				
Pretest Quintile:	Bottom	2nd	3rd	4th	Top	Bottom	2nd	3rd	4th	Top
Control (C) Mean	-0.734	-0.577	-0.432	-0.0891	0.640	-0.570	-0.574	-0.467	-0.0571	0.695
(Std. Dev.)	(0.289)	(0.402)	(0.549)	(0.685)	(0.935)	(0.536)	(0.452)	(0.608)	(0.781)	(1.051)
A. Treatment Variables										
Incentives Only (G1)	0.172**	0.0382	0.0258	-0.0849	-0.165	0.0726	0.0884	0.200**	-0.104	-0.0338
	(0.0858)	(0.0861)	(0.0878)	(0.114)	(0.145)	(0.0776)	(0.0698)	(0.101)	(0.112)	(0.151)
	[0.0490]	[0.973]	[1]	[0.443]	[0.207]	[0.545]	[0.262]	[0.0530]	[0.309]	[1]
Technology Only (G2)	-0.00467	0.0253	-0.0377	-0.0768	0.212	-0.0507	0.0297	-0.0286	-0.211*	-0.0357
	(0.0514)	(0.0760)	(0.0631)	(0.104)	(0.176)	(0.0704)	(0.0724)	(0.0670)	(0.111)	(0.165)
	[0.448]	[0.973]	[1]	[0.443]	[0.207]	[0.545]	[0.451]	[0.288]	[0.214]	[1]
Both (G3)	0.246***	0.208**	0.320***	0.286**	0.404***	0.140	0.236***	0.342***	0.160	0.273*
	(0.0823)	(0.0842)	(0.0862)	(0.116)	(0.150)	(0.0887)	(0.0805)	(0.0981)	(0.117)	(0.156)
	[0.0100]	[0.0460]	[0.00100]	[0.0450]	[0.0250]	[0.545]	[0.0120]	[0.00200]	[0.214]	[0.325]
B. Linear Combinations of Estimators, Other Tests and Details										
β_{T3} - β_{T1}	0.0738	0.170**	0.294***	0.371***	0.569***	0.0670	0.148	0.142	0.265**	0.307**
	(0.0946)	(0.0826)	(0.0955)	(0.104)	(0.169)	(0.0923)	(0.0957)	(0.124)	(0.104)	(0.138)
β_{T3} - β_{T2}	0.251***	0.183**	0.357***	0.363***	0.192	0.190**	0.207**	0.371***	0.371***	0.309*
	(0.0768)	(0.0849)	(0.0839)	(0.104)	(0.205)	(0.0867)	(0.0952)	(0.0904)	(0.103)	(0.159)
β_{T3} - β_{T1} - β_{T2}	0.0785	0.145	0.332***	0.448***	0.357	0.118	0.118	0.170	0.476***	0.343
	(0.114)	(0.117)	(0.121)	(0.148)	(0.253)	(0.117)	(0.120)	(0.141)	(0.154)	(0.220)
Observations	4,079	4,079	4,079	4,079	4,079	3,919	3,919	3,919	3,919	3,919
R-squared	0.492	0.492	0.492	0.492	0.492	0.469	0.469	0.469	0.469	0.469
Clusters	170	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.00260	0.0640	0.000400	0.00140	0.00430	0.116	0.0279	0.000100	0.00410	0.111

Note : Difference-in-means coefficients. Controls are as in columns (3), (6) and (9) of Table 3: age, year-0 score, randomization-block (five-region) indicators, and a (3rd-degree) polynomial of attrition-propensity score. Standard errors: clustered by school in parentheses. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. Third row in brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses) as described in Anderson (2008).

Table A6: Unequal Piece Rates in Year 1: Effects on Outcomes and Perceptions on Fairness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Year 1 (Piece rates were unequal)					Year 2 (Piece rates were equal)			
	Present	Hrs/wk of Math Study	Year 1 Z-score	Subsidy was Motivating: "Always"	Unequal Piece-rates were "Demotivating"	Present	Hrs/wk of Math Study	Year 2 Z-score	Year 1 Piece-rates were "Unfair"
A.Incentives Only (G1) Constant Term and Incremental Effects of Piece Rates (Base Group = "Promised $0.125 / Mark")									
Incentives Only (G1)	0.0287	0.401	0.0418	0.774***	0.0903***	0.100***	-0.00783	-0.105	0.181***
	(0.0273)	(0.425)	(0.0880)	(0.0388)	(0.0155)	(0.0323)	(0.676)	(0.0830)	(0.0420)
Incentives Only (G1) x Promised $0.25 / Mark	0.0261	0.0474	-0.00486	0.0595**	-0.00212	-0.0228	-0.364	0.116**	0.00292
	(0.0263)	(0.225)	(0.0366)	(0.0290)	(0.0185)	(0.0320)	(0.466)	(0.0457)	(0.0225)
Incentives Only (G1) x Promised $0.5 / Mark	-0.0232	-0.176	0.0450	0.0693**	-0.0240	-0.0287	-0.941*	0.0451	-0.0406
	(0.0265)	(0.222)	(0.0451)	(0.0327)	(0.0209)	(0.0303)	(0.500)	(0.0464)	(0.0275)
Incentives Only (G1) x Promised $0.75 / Mark	0.0265	-0.184	-0.00741	0.0753**	-0.0323**	-0.0116	-0.351	0.0871	-0.0478*
	(0.0242)	(0.237)	(0.0543)	(0.0331)	(0.0160)	(0.0288)	(0.502)	(0.0604)	(0.0288)
B. Both (G3) Constant Term and Incremental Effects of Incremental Piece Rates (Base Group = "Promised $0.125 / Mark")									
Both (G3)	0.0426	0.746	0.0586	0.779***	0.0752***	0.0425	1.198*	0.197**	0.231***
	(0.0261)	(0.466)	(0.102)	(0.0585)	(0.0171)	(0.0301)	(0.643)	(0.0916)	(0.0488)
Both (G3) x Promised $0.25 / Mark	-0.00170	0.363	0.0732*	0.0129	-0.0161	0.0237	-0.353	0.0878	-0.0207
	(0.0266)	(0.278)	(0.0397)	(0.0298)	(0.0188)	(0.0338)	(0.507)	(0.0652)	(0.0400)
Both (G3) x Promised $0.5 / Mark	0.0106	0.194	0.131**	0.00243	-0.00366	-0.00717	0.100	0.133*	-0.0546*
	(0.0230)	(0.284)	(0.0533)	(0.0258)	(0.0208)	(0.0296)	(0.562)	(0.0705)	(0.0316)
Both (G3) x Promised $0.75 / Mark	0.0156	0.105	0.0715	0.00867	-0.00930	0.0645**	0.238	0.122*	-0.0595
	(0.0268)	(0.271)	(0.0558)	(0.0272)	(0.0164)	(0.0293)	(0.663)	(0.0641)	(0.0368)
Selection and Other Ctrls.	X	X	X	X	X	X	X	X	X
Observations	5,508	4,697	4,697	4,697	4,697	5,508	4,079	4,079	4,079
R-squared	0.036	0.103	0.552	0.673	0.038	0.080	0.062	0.487	0.120
Clusters	170	170	170	170	170	170	170	170	170
Pr. > Joint F, All Incremental Effects = 0	0.532	0.800	0.206	0.320	0.204	0.155	0.393	0.0978	0.210

Note : Difference-in-means coefficients. Observations: year 1 (Oct. '16) and year 2 (Oct. '17) survey responses. Controls include age, year-0 score, randomization-block (five-region) indicators, and a (3rd-degree) polynomial of selection-propensity score (probit analogues of columns (1) and (3) in Table 2). Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. Third row in brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses) as described in Anderson (2008).

Table A7: Breakdown of Subtopics Evaluated on Year 3 Mock and O Level tests

	(1)		(2)	
Test Type (Timing):	Mock (Oct. '18)		Real (Nov. '18)	
Statistic:	Total Marks		Total Marks	
A. Shared Subtopics, Shared Marks				
Form 1, Chapter 2, Fractions, Decimals and Percentages	1.5		1.5	
Form 1, Chapter 9, Ratios, Profit and Loss	3		3	
Form 2, Chapter 1, Exponents and Radicals	1.5		1.5	
Form 2, Chapter 2, Algebra	7		7	
Form 2, Chapter 3, Quadratic Equations	3		3	
Form 2, Chapter 4, Logarithms	1.5		1.5	
Form 2, Chapter 8, Pythagoras Theorem	3		3	
Form 2, Chapter 9, Set Theory	3		3	
Form 2, Chapter 10, Statistics	5		5	
Form 3, Chapter 2, Functions	7		7	
Form 3, Chapter 4, Rates and Variations	5		5	
Form 3, Chapter 5, Sequence and Series	6		6	
Form 3, Chapter 6, Circles	1.5		1.5	
Form 3, Chapter 8, Accounts	10		10	
Form 4, Chapter 1, Coordinate Geometry	3		3	
Form 4, Chapter 2, Areas and Perimeters	3		3	
Form 4, Chapter 4, Probability	5		5	
Form 4, Chapter 5, Trigonometry	3		3	
Form 4, Chapter 6, Vectors	3		3	
Form 4, Chapter 8, Linear Programming	5		5	
(Subtotal Marks of Panel A)		(80)		(80)
B. Shared Subtopics, Different Marks				
Form 2, Chapter 1, Exponents and Radicals	1.5			
Form 2, Chapter 2, Algebra	2			
Form 2, Chapter 3, Quadratic Equations	1.5			
Form 2, Chapter 4, Logarithms	1.5			
Form 2, Chapter 10, Statistics			5	
Form 3, Chapter 2, Functions	4			
Form 3, Chapter 4, Rates and Variations			2	
Form 4, Chapter 2, Areas and Perimeters			3	
Form 3, Chapter 6, Circles			0.5	
Form 4, Chapter 8, Linear Programming			5	
(Subtotal Marks of Panel B)		(10.5)		(15.5)
C. Different Subtopics, Different Marks				
Form 1, Chapter 1, Numbers			3	
Form 1, Chapter 10, Real Numbers			1.5	
Form 1, Chapter 4, Approximations	3			
Form 2, Chapter 5, Congruence and Similarity	1.5			
Form 3, Chapter 7, The Earth as a Sphere	5			
(Subtotal Marks of Panel C)		(9.5)		(4.5)

Note: Breakdown of topics covered on year 3 program mock test (Oct. '18) and O Level mathematics certification test (Nov. '18).

Table A8: Structural Parameter Estimates: Benchmark Model

#	Value	Std. Err.	Name	Label
1	1.08	(0.066)	α_0	student preparedness (K_0) exp.
2	0.140	(0.041)	α_1	student effort (E) exp.
3	0.163	(0.036)	γ_0	teacher ability (A) exp.
4	0.0011	(0.0027)	γ_1	teacher effort (R) exp.
5	0.148	(0.053)	δ_{cons}	regression rate (δ) cons.
6	0.00012	(0.00014)	$\Delta\delta_{K0}$	δ (K_{i0} - K_{min}) coef.
7	0.046	(0.0071)	τ_{cons}	factor productivity (τ) cons.
8	0.0068	(0.0015)	τ_{SMR}	τ SMR tech. coef.
9	4.79	(8.41)	$\pi_{\$}$	knowl. pref. (π) SMR \$ coef.
10	0.900	(0.569)	π_{scale}	π measurement scale coef.
11	1350.0	(725.1)	θ_{cons}	promotion value (θ) cons.
12	54.4	(282.5)	θ_{STEM}	θ science-intended coef.
13	1E-08	(26.8)	ι	test-attendance effort cons.
14	1.41	(2.29)	$\underline{E}$	E location
15	2.58	(0.512)	p	E exp.
16	-	(-)	$\underline{K}_0$	minimum K_0 threshold ($\underline{K}_0$)
17	-	(-)	$\Delta\underline{K}_{0,SMR}$	$\underline{K}_0$ SMR tech. coef.
18	-	(-)	$\underline{\pi}$	mininum π threshold ($\underline{\pi}$)
19	8.59	(5.57)	β_0^{K0}	K_0 deq. cons.
20	0.171	(1.70)	β_1^{K0}	K_0 deq. coef.--female
21	0.476	(0.056)	β_2^{K0}	"--Y_0 math score
22	1.07	(0.837)	β_3^{K0}	"--age
23	0.444	(0.063)	β_4^{K0}	"--FTNA nonmath average
24	-0.907	(2.67)	β_5^{K0}	"--parental education
25	4.47	(0.842)	β_0^{π}	π deq. cons.
26	0.094	(0.442)	β_1^{π}	π deq. coef.--female
27	0.0019	(0.00080)	β_2^{π}	"--FTNA nonmath average
28	0.079	(0.377)	β_3^{π}	"--parental education
29	0.191	(0.457)	β_4^{π}	"--intended occupation = STEM
30	0.660	(0.131)	β_0^{A}	A deq. cons.
31	0.041	(0.036)	β_1^{A}	A deq. coef.--teaches math fulltime
32	0.037	(0.027)	β_2^{A}	"--has bachelor's degree
33	0.012	(0.0083)	β_3^{A}	"--# years taught math
34	-0.00050	(0.00042)	β_4^{A}	"--# years taught math^2
35	0.464	(0.282)	β_0^{R}	R deq. cons.
36	-0.00015	(0.028)	β_1^{R}	R deq. coef.--total teacher hrs/wk
37	0.021	(0.215)	β_2^{R}	"--teaches math fulltime
38	0.020	(0.212)	β_3^{R}	"--has bachelor's degree
39	0.012	(0.042)	β_4^{R}	"--# years taught math
40	-0.00032	(0.00070)	β_5^{R}	"- # years taught math^2
41	-2.40	(6.96)	β_0^{ζ}	test-attendance cost (ζ) deq. cons.
42	-0.0094	(0.026)	β_1^{ζ}	ζ deq. coef.--util. diff.
43	-0.113	(0.734)	β_2^{ζ}	"--female
44	0.088	(0.426)	β_3^{ζ}	"--age
45	-0.119	(1.01)	β_4^{ζ}	"--parental education
46	0.016	(0.130)	β_5^{ζ}	"--commute distance
47	29.2	(1.77)	$\sigma(\mu_j^{K0})$	std. dev. school effect
48	0.770	(0.179)	$\sigma(\mu_j^{\pi})$	"
49	0.226	(0.015)	$\sigma(\mu_j^{A})$	"
50	0.209	(0.205)	$\sigma(\mu_j^{R})$	"
51	0.096	(0.256)	$\rho(\mu_j^{K0}, \mu_j^{\pi})$	corr. coef. school effect
52	0.0062	(0.024)	$\rho(\mu_j^{K0}, \mu_j^{A})$	"
53	0.9999	(0.016)	$\rho(\mu_j^{A}, \mu_j^{R})$	"
54	15.0	(4.36)	$\sigma(\omega_{ij}^{K0})$	std. dev. indiv. effect
55	0.0088	(0.025)	$\sigma(\omega_{ij}^{\pi})$	"
56	0.9999	(0.190)	$\rho(\omega_{ij}^{K0}, \omega_{ij}^{\pi})$	corr. coef. indiv. effect
57	2995.6	(641.4)	$\sigma^2(\epsilon_{ij}^{K0, m1})$	K_0 meq. 1 (Y_0 math score) var.
58	0.017	(0.0060)	$\beta_{slope}^{K0, m2}$	K_0 meq. 2 (FTNA math)--slope
59	10.5	(3.43)	$\beta_{cut1}^{K0, m2}$	"--cut1
60	1.33	(0.994)	$\beta_{diff2}^{K0, m2}$	"--cut2 - cut1
61	1.56	(1.80)	$\beta_{diff3}^{K0, m2}$	"--cut3 - cut2
62	1.03	(3.04)	$\beta_{diff4}^{K0, m2}$	"--cut4 - cut3
63	7.94	(2.17)	$\sigma^2(\epsilon_{ij}^{\pi, m1})$	π meq. 1 (Y_0 nonmath hrs) var.
64	0.631	(0.382)	$\beta_{slope}^{\pi, m2}$	π meq. 2 (likes math)--slope
65	1.21	(2.41)	$\beta_{cut1}^{\pi, m2}$	"--cut1
66	0.380	(0.664)	$\beta_{diff2}^{\pi, m2}$	"--cut2-cut1
67	1.58	(0.499)	$\beta_{diff3}^{\pi, m2}$	"--cut3-cut2
68	0.012	(0.031)	$\sigma^2(\epsilon_{ij}^{A, m1})$	A meq. 1 (teacher knowl.) var.
69	-0.0019	(0.810)	$\beta_0^{A, m2}$	A meq. 2 (teacher control)--cons.
70	0.893	(1.15)	$\beta_1^{A, m2}$	"--slope
71	0.014	(0.035)	$\sigma^2(\epsilon_{ij}^{A, m2})$	"--var.
72	0.013	(0.026)	$\sigma^2(\epsilon_{ij}^{R, m1})$	R meq. 1 (teacher cares) var.
73	0.137	(0.909)	$\beta_0^{R, m2}$	R meq. 2 (teacher attends)--cons.
74	0.915	(1.66)	$\beta_1^{R, m2}$	"--slope
75	0.011	(0.031)	$\sigma^2(\epsilon_{ij}^{R, m2})$	"--var.
76	39.7	(5.98)	$\sigma^2(\epsilon_{ij}^{E, m1})$	E meq. 1 (Y_2 math hrs) var.
77	0.326	(0.273)	$\beta_{slope}^{E, m2}$	E meq. 2 (Y_2 math attention)--cons.
78	0.356	(1.77)	$\beta_{cut1}^{E, m2}$	"--cut1
79	0.955	(0.428)	$\beta_{diff2}^{E, m2}$	"--cut2-cut1
80	0.947	(0.348)	$\beta_{diff3}^{E, m2}$	"--cut3-cut2
81	1.02	(0.367)	$\beta_{diff4}^{E, m2}$	"--cut4-cut3
82	3614.2	(868.1)	$\sigma^2(\epsilon_{ij}^{Ki, m1})$	K_i meq. 1 (Y_2 math score) var.

Notes: "exp" stands for exponent; "cons," constant; "coef," coefficient"; "knowl. pref," student knowledge preference; "deq," latent-factor-determinant equation; "meq," measurement-errror equation; "var," variance.

Table A9: Structural Parameter Estimates: Productivity-Thresholds Model

#	Value	Std. Err.	Name	Label
1	1.08	(0.058)	α_0	student preparedness (K_0) exp.
2	0.129	(0.065)	α_1	student effort (E) exp.
3	0.0010	(0.0064)	γ_0	teacher ability (A) exp.
4	0.0021	(0.0086)	γ_1	teacher effort (R) exp.
5	0.020	(0.027)	δ_{cons}	regression rate (δ) cons.
6	0.00033	(0.00019)	$\Delta\delta_{K0}$	δ (K_{i0} - K_{min}) coef.
7	0.044	(0.014)	τ_{cons}	factor productivity (τ) cons.
8	0.015	(0.0037)	τ_{SMR}	τ SMR tech. coef.
9	14.1	(6.75)	$\pi_{\$}$	knowl. pref. (π) SMR \$ coef.
10	1.00	(0.151)	π_{scale}	π measurement scale coef.
11	395.7	(78.3)	θ_{cons}	promotion value (θ) cons.
12	-33.7	(155.9)	θ_{STEM}	θ science-intended coef.
13	3.11	(1.15)	ι	test-attendance effort cons.
14	5.61	(1.40)	$\underline{E}$	E location
15	3.58	(1.56)	p	E exp.
16	554.1	(48.8)	$\underline{K}_0$	minimum K_0 threshold ($\underline{K}_0$)
17	-154.9	(271.7)	$\Delta\underline{K}_{0,SMR}$	$\underline{K}_0$ SMR tech. coef.
18	7.89	(0.736)	$\underline{\pi}$	mininum π threshold ($\underline{\pi}$)
19	8.39	(5.09)	β_0^{K0}	K_0 deq. cons.
20	-2.18	(9.35)	β_1^{K0}	K_0 deq. coef.--female
21	0.476	(0.049)	β_2^{K0}	"--Y_0 math score
22	1.32	(0.523)	β_3^{K0}	"--age
23	0.444	(0.049)	β_4^{K0}	"--FTNA nonmath average
24	-3.37	(14.6)	β_5^{K0}	"--parental education
25	4.67	(0.508)	β_0^{π}	π deq. cons.
26	0.087	(0.072)	β_1^{π}	π deq. coef.--female
27	0.0018	(0.00087)	β_2^{π}	"--FTNA nonmath average
28	0.085	(0.574)	β_3^{π}	"--parental education
29	0.096	(0.546)	β_4^{π}	"--intended occupation = STEM
30	0.616	(0.465)	β_0^{A}	A deq. cons.
31	0.072	(0.408)	β_1^{A}	A deq. coef.--teaches math fulltime
32	-0.015	(0.334)	β_2^{A}	"--has bachelor's degree
33	0.0057	(0.076)	β_3^{A}	"--# years taught math
34	0.00010	(0.0030)	β_4^{A}	"--# years taught math^2
35	0.430	(0.423)	β_0^{R}	R deq. cons.
36	-0.0060	(0.034)	β_1^{R}	R deq. coef.--total teacher hrs/wk
37	0.042	(0.395)	β_2^{R}	"--teaches math fulltime
38	-0.0067	(0.294)	β_3^{R}	"--has bachelor's degree
39	0.016	(0.070)	β_4^{R}	"--# years taught math
40	-0.00020	(0.0031)	β_5^{R}	"- # years taught math^2
41	-2.40	(6.13)	β_0^{ζ}	test-attendance cost (ζ) deq. cons.
42	-0.0016	(0.0054)	β_1^{ζ}	ζ deq. coef.--util. diff.
43	-0.114	(0.820)	β_2^{ζ}	"--female
44	0.086	(0.379)	β_3^{ζ}	"--age
45	-0.102	(1.06)	β_4^{ζ}	"--parental education
46	0.017	(0.128)	β_5^{ζ}	"--commute distance
47	32.0	(5.64)	$\sigma(\mu_j^{K0})$	std. dev. school effect
48	0.682	(0.330)	$\sigma(\mu_j^{\pi})$	"
49	0.212	(0.238)	$\sigma(\mu_j^{A})$	"
50	0.205	(0.201)	$\sigma(\mu_j^{R})$	"
51	0.182	(0.419)	$\rho(\mu_j^{K0}, \mu_j^{\pi})$	corr. coef. school effect
52	-0.014	(0.249)	$\rho(\mu_j^{K0}, \mu_j^{A})$	"
53	1	(0.0033)	$\rho(\mu_j^{A}, \mu_j^{R})$	"
54	11.3	(4.43)	$\sigma(\omega_{ij}^{K0})$	std. dev. indiv. effect
55	0.188	(0.113)	$\sigma(\omega_{ij}^{\pi})$	"
56	1	(0.044)	$\rho(\omega_{ij}^{K0}, \omega_{ij}^{\pi})$	corr. coef. indiv. effect
57	3042.3	(671.5)	$\sigma^2(\epsilon_{ij}^{K0, m1})$	K_0 meq. 1 (Y_0 math score) var.
58	0.017	(0.0066)	$\beta_{slope}^{K0, m2}$	K_0 meq. 2 (FTNA math)--slope
59	10.5	(3.82)	$\beta_{cut1}^{K0, m2}$	"--cut1
60	1.33	(1.14)	$\beta_{diff2}^{K0, m2}$	"--cut2 - cut1
61	1.56	(1.88)	$\beta_{diff3}^{K0, m2}$	"--cut3 - cut2
62	1.03	(2.98)	$\beta_{diff4}^{K0, m2}$	"--cut4 - cut3
63	8.04	(2.25)	$\sigma^2(\epsilon_{ij}^{\pi, m1})$	π meq. 1 (Y_0 nonmath hrs) var.
64	0.631	(0.449)	$\beta_{slope}^{\pi, m2}$	π meq. 2 (likes math)--slope
65	1.27	(2.96)	$\beta_{cut1}^{\pi, m2}$	"--cut1
66	0.383	(0.698)	$\beta_{diff2}^{\pi, m2}$	"--cut2-cut1
67	1.56	(0.493)	$\beta_{diff3}^{\pi, m2}$	"--cut3-cut2
68	0.013	(0.032)	$\sigma^2(\epsilon_{ij}^{A, m1})$	A meq. 1 (teacher knowl.) var.
69	-0.015	(1.06)	$\beta_0^{A, m2}$	A meq. 2 (teacher control)--cons.
70	0.915	(1.47)	$\beta_1^{A, m2}$	"--slope
71	0.014	(0.030)	$\sigma^2(\epsilon_{ij}^{A, m2})$	"--var.
72	0.013	(0.025)	$\sigma^2(\epsilon_{ij}^{R, m1})$	R meq. 1 (teacher cares) var.
73	0.138	(0.753)	$\beta_0^{R, m2}$	R meq. 2 (teacher attends)--cons.
74	0.912	(1.39)	$\beta_1^{R, m2}$	"--slope
75	0.010	(0.030)	$\sigma^2(\epsilon_{ij}^{R, m2})$	"--var.
76	39.5	(7.72)	$\sigma^2(\epsilon_{ij}^{E, m1})$	E meq. 1 (Y_2 math hrs) var.
77	0.297	(0.247)	$\beta_{slope}^{E, m2}$	E meq. 2 (Y_2 math attention)--cons.
78	0.202	(1.63)	$\beta_{cut1}^{E, m2}$	"--cut1
79	0.940	(0.430)	$\beta_{diff2}^{E, m2}$	"--cut2-cut1
80	0.933	(0.350)	$\beta_{diff3}^{E, m2}$	"--cut3-cut2
81	1.01	(0.361)	$\beta_{diff4}^{E, m2}$	"--cut4-cut3
82	3708.0	(898.1)	$\sigma^2(\epsilon_{ij}^{Ki, m1})$	K_i meq. 1 (Y_2 math score) var.

Notes: "exp" stands for exponent; "cons," constant; "coef," coefficient"; "knowl. pref," student knowledge preference; "deq," latent-factor-determinant equation; "meq," measurement-errror equation; "var," variance.

Table A10: Effects on O Level Mathematics Certification Examination Results and Breakdown of Difference from Mock Test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Unit:	Z-score Converted from Grade Brackets				Pass Indicator			
Variable Type:	Nov. '18 O Level	Oct. '18 Mock Test	Improvement on O Level	Deterioation on O Level	Nov. '18 O Level	Oct. '18 Mock Test	Improvement on O Level	Deterioation on O Level
Control (C) Mean	-0.0480	-0.0975	0.114	-0.0649	0.0962	0.0651	0.0401	-0.00900
(Std. Dev.)	(0.880)	(0.784)	(0.521)	(0.272)	(0.295)	(0.247)	(0.196)	(0.0945)
A. Treatment Variables								
Incentives Only (G1)	-0.0174	0.0215	-0.0195	-0.0194	-0.00755	0.00960	-0.00793	-0.00921
	(0.0495)	(0.0510)	(0.0257)	(0.0229)	(0.0161)	(0.0149)	(0.00929)	(0.00714)
	[1]	[0.816]	[0.818]	[0.362]	[1]	[1]	[0.652]	[0.178]
Technology Only (G2)	0.00332	0.0381	-0.00496	-0.0298	-0.0118	-0.00162	-0.00246	-0.00770
	(0.0494)	(0.0788)	(0.0282)	(0.0279)	(0.0132)	(0.0169)	(0.0109)	(0.00633)
	[1]	[0.816]	[1]	[0.362]	[1]	[1]	[1]	[0.178]
Both (G3)	0.0245	0.166***	-0.0495**	-0.0921***	0.000843	0.0483***	-0.0203**	-0.0272***
	(0.0443)	(0.0617)	(0.0225)	(0.0318)	(0.0136)	(0.0182)	(0.00842)	(0.00981)
	[1]	[0.0250]	[0.0960]	[0.0140]	[1]	[0.0270]	[0.0550]	[0.0190]
C. Linear Combinations of Estimators, Other Tests and Details								
$\beta_{T3} - \beta_{T1}$	0.0419	0.145**	-0.0300	-0.0727**	0.00839	0.0387**	-0.0123	-0.0180
	(0.0508)	(0.0612)	(0.0218)	(0.0363)	(0.0159)	(0.0187)	(0.00755)	(0.0110)
$\beta_{T3} - \beta_{T2}$	0.0212	0.128	-0.0445*	-0.0623	0.0126	0.0500**	-0.0178*	-0.0195*
	(0.0512)	(0.0898)	(0.0254)	(0.0409)	(0.0124)	(0.0206)	(0.00981)	(0.0108)
$\beta_{T3} - \beta_{T1} - \beta_{T2}$	0.0386	0.106	-0.0250	-0.0429	0.0202	0.0404	-0.00986	-0.0103
	(0.0708)	(0.100)	(0.0356)	(0.0463)	(0.0206)	(0.0255)	(0.0132)	(0.0130)
Observations	5,965	5,965	5,965	5,965	5,965	5,965	5,965	5,965
R-squared (Mean)	0.427	0.320	0.113	0.0385	0.378	0.277	0.0666	0.0139
Clusters	170	170	170	170	170	170	170	170
Pr. > Joint F, All Treat. = 0	0.869	0.0523	0.101	0.0295	0.725	0.0488	0.0594	0.0370

Note: Difference-in-means coefficients. Columns (1)-(4) concern Z-scores converted from the midpoint of each student's grade bracket: 88 (A), 67 (B), 52 (C), 37 (D) and 15 (F). Columns (5)-(8) concern pass indicators given by a grade of D or above. Controls: age, year-0 score, commute distance. Regressions drop transferred students. Controls were missing at random for about 10% of students; estimates were obtained using multiple impuation and combined using Rubin's (1987) formulas. Standard errors: clustered by school. Levels of significance: *** p<0.01, ** p<0.05, * p<0.10. In brackets: Benjamini-Krieger-Yekutieli (2006) sharpened two-stage q-values (for three hypotheses).

B Appendix Figures

Figure B1: Experimental Design

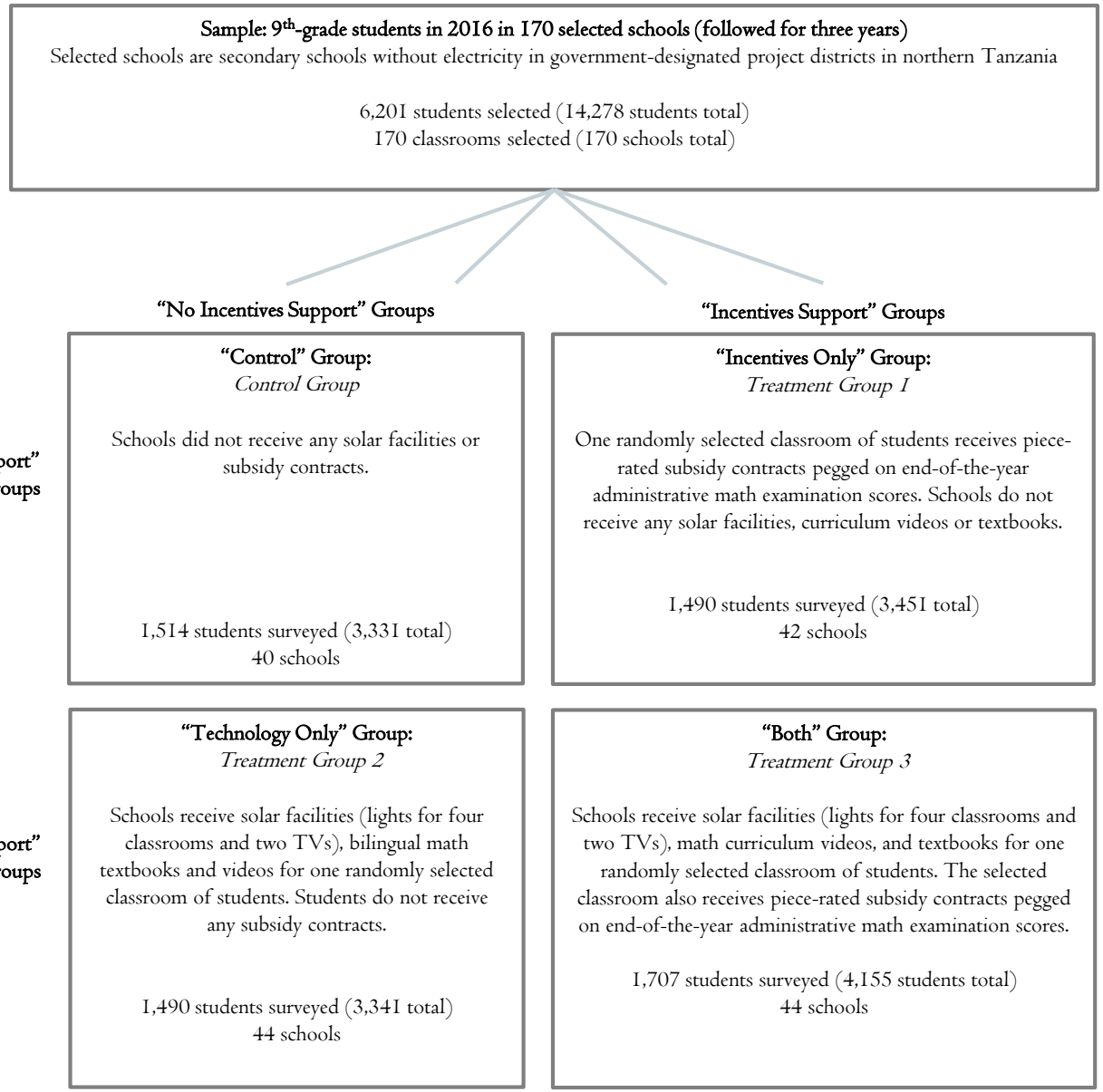


Figure B2: Experimental Timeline

September 2015

- Pilot Phase Solar Installations: Technology Support groups randomly selected and pilot systems tested (lights for 1~2 classrooms and one 16" TV)

November 2015

- Grant awarded for full study
- Classrooms sampled and cross-randomized into Incentives groups

February 2016

- Pre-incentives Examination & Survey: Mock test for 8th-grade students administered
- Year 1 Incentives Contracts and Technology Support Delivered: Contracts announced and books & videos delivered to respective support groups

May 2016

- Main Phase Solar Installations: Additional solar systems (lights for 1~2 additional classrooms and one 19" TV) installed for Technology Support groups

October 2016

- Year 1 Mock Examination & Survey: Administered for 9th-grade students
- Incentive awards distributed

November 2016

- '16 National Promotional Examinations: administered for 9th-grade students

February 2017

- Year 2 Incentives Contracts and Technology Support Delivered: Contracts announced and books & videos delivered to respective support groups

October 2017

- Year 2 Mock Examination & Survey: Administered for 10th-grade students
- Incentive awards distributed

December 2017

- Solar Installations for Remaining Schools: Solar systems installed in Technology Support groups also installed in remaining groups, as promised. Only the original Technology Support groups continue receiving new textbooks and videos.

February 2018

- Year 3 Incentives Contracts and Technology Support Delivered: Contracts announced and books & videos delivered to respective support groups

October 2018

- Year 3 Mock Examination & Survey: Administered for 11th-grade students
- Incentive awards distributed

November 2018

- Year 3 Certification Examination: Administered for 11th-grade students

Figure B3: Textbooks, Videos and Final Year Tests

Period	Learning Material	URL
Year 1 (Form 2) [Delivered Feb., 2016]	Sharpening Mathematics Review Textbook	https://tinyurl.com/temp-year1-yssm-textbook
	Sharpening Mathematics Review Videos	https://tinyurl.com/year1-yssm-video
Year 2 (Form 3) [Delivered Feb., 2017]	Sharpening Mathematics Review Textbook	https://tinyurl.com/temp-year2-yssm-textbook
	Sharpening Mathematics Review Videos	https://tinyurl.com/year2-yssm-video
Year 3 (Form 4) [Delivered Feb., 2018]	Sharpening Mathematics Review Textbook	https://tinyurl.com/temp-year3-yssm-textbook
	Sharpening Mathematics Review Videos	https://tinyurl.com/year3-yssm-video
End of Year 3 (Form 4) [Administered Oct., 2018]	Sharpening Mathematics Review (Mock) Test	https://tinyurl.com/year-3-mock-test
End of Year 3 (Form 4) [Administered Nov., 2018]	Certificate of Secondary Education Examination – Basic Mathematics (Real) Test	https://tinyurl.com/year3-real-test

C Appendix Proof

Proposition C.1. *The optimization problem defined by eq.* (4) *with a single pass cutoff—that is, marginal benefit* $MB(K \mid \pi, \theta, T, \sigma) = \pi + \frac{\theta}{\sigma}\phi\left(\frac{K-T}{\sigma}\right)$ *and marginal cost* $MC(K \mid a, b, \lambda) = a(K-b)^{\lambda}$*—has one unique argument maximum* K^* *in the general case. The problem has dual maxima only in a degenerate case: a case of measure zero when latent factors a and b are being drawn from a continuous distribution. The same result holds for the modified problem with* $MB = \pi + \frac{\theta}{\sigma}\phi\left(\frac{\min\{K-T,0\}}{\sigma}\right)$.

Proof First, note that the problem has at least one non-degenerate solution, given that $MB \geq \pi > 0$ and $MB \to \pi$ in the limits, and that MC rises from 0 (at $K = b$) and tends to infinity (as $K \to \infty$).

Next, I show that the logarithmic derivatives of $\tilde{MB} = MB - \pi$ and $\tilde{MC} = MC - \pi$ intersect in at most two points. The logarithmic derivative of $\tilde{MB}$ falls linearly in K.[59] The logarithmic derivative of $\tilde{MC}$ falls from infinity (beginning at $K = b$ where $\tilde{MC} = 0$), and is strictly convexly in K.[60] Hence, these two logarithmic derivatives intersect in at most two points, since, by definition, a strictly convex curve intersects any line in at most two points.

Next, I show that $\tilde{MB}$ and $\tilde{MC}$ intersect in at most three points; this is because each interval of the set $K \in [b, \infty)$ partitioned by the points where the logarithmic derivatives are equal admits at most one point where $\tilde{MB}$ intersects $\tilde{MC}$. To see this, assume for contradiction that $\tilde{MB}$ intersects $\tilde{MC}$ more than once in any such interval I. Let $K_1^X < \ldots < K_N^X$ denote the intersections in I. Since either $\frac{\tilde{MB}'}{\tilde{MB}} > \frac{\tilde{MC}'}{\tilde{MC}}, \forall K \in I$, or $\frac{\tilde{MB}'}{\tilde{MB}} < \frac{\tilde{MC}'}{\tilde{MC}}, \forall K \in I$, and the denominators are equal at intersections (K_n^X), it must be that either $\tilde{MB}' > \tilde{MC}', \forall K_n^X \in I$, or $\tilde{MB}' < \tilde{MC}', \forall K_n^X \in I$. This means that the continuous curve $f(K) = \tilde{MB} - \tilde{MC}$ (with $f'(K) = \tilde{MB}' - \tilde{MC}'$) has two consecutive roots, K_1^X and K_2^X, where the function f crosses 0 with slopes of the same sign, a

[59] $\frac{\tilde{MB}'}{\tilde{MB}} = \frac{T-K}{\sigma^2}$.

[60] $\frac{\tilde{MC}'}{\tilde{MC}} = \frac{\lambda}{(K-b)-\pi(K-b)^{1-\lambda}/a}$. Note that the denominator is positive where $MC > \pi$. Hence, $\forall K$ s.t. $MC > \pi$: $\left(\frac{\tilde{MC}'}{\tilde{MC}}\right)'' = \frac{2\lambda\left(1+(\lambda-1)\pi(K-b)^{-\lambda}/a\right)}{\left[(K-b)-\pi(K-b)^{1-\lambda}/a\right]^3} + \frac{\lambda^2(\lambda-1)\pi(K-b)^{-1-\lambda}/a}{\left[(K-b)-\pi(K-b)^{1-\lambda}/a\right]^2} > 0$.

contradiction.[61]

Hence, MB and MC, which are $\tilde{MB}$ and $\tilde{MC}$ just shifted up by a constant, intersect in at most three points.

When MB and MC intersect once, the argmax K^* is unique, since MC crosses MB from below. When MB and MC intersect twice, the argmax K^* is again unique, since the area between the two intersections (bounded vertically by the MB and MC curves) is always strictly positive or strictly negative. When MB and MC intersect three times, let $K_{p1} < K_{p2} < K_{p3}$ denote the three intersections. K_{p2} immediately follows an interval in which MB remains below MC; hence, K_{p2} cannot be an argmax, leaving only the possibilities of a unique or dual argmax.

Finally, I show that dual maxima occur only in a degenerate case. Let $U(K) = \int MB - MC \, \mathrm{d}K$. Fixing parameters $\{\pi, \theta, T, \sigma, \lambda\}$ and factor b, the number of values of a that satisfies $D(a, K_{p1}, K_{p3}) = U(K_{p3}, a) - U(K_{p1}, a) = 0$ is at most one. This is because, by the envelope theorem, $\frac{\mathrm{d}D(a,K_{p1}(a),K_{p3}(a))}{\mathrm{d}a} = \frac{\partial D(a,K_{p1},K_{p3})}{\partial a} = -\frac{1}{\lambda+1}(K_{p3}-b)^{\lambda+1} + \frac{1}{\lambda+1}(K_{p1}-b)^{\lambda+1} < 0$; because $D(a, K_{p1}, K_{p1})$ is always decreasing in a, it can cross 0 at no more than one value of a. Likewise, $\frac{\mathrm{d}D(b,K_{p1}(b),K_{p3}(b))}{\mathrm{d}b} = \frac{\partial D(b,K_{p1},K_{p3})}{\partial b} = a(K_{p3}-b)^{\lambda} - a(K_{p1}-b)^{\lambda} > 0$; because $D(b, K_{p1}, K_{p3})$ is always increasing in b, it can cross 0 at no more than one value of b. Therefore, dual maxima can occur only given a degenerate set (a, b): a case of measure 0 when (a, b) are being drawn from a continuous distribution.

The same result holds for the modified problem with $MB = \pi + \frac{\theta}{\sigma}\phi\left(\frac{\min\{K-T,0\}}{\sigma}\right)$. Note that in the original problem, MB and MC can intersect at most once where $K \geq T$, since MB is strictly decreasing in K while MC is strictly increasing where $K \geq T$. Likewise in the modified problem, MB and MC can intersect at most once where $K \geq T$, since MB is flat while MC is strictly increasing where $K \geq T$. The shape of the problem and hence the solutions remain identical where $K < T$. ∎

[61] By the intermediate value theorem, given two consecutive roots of a continuous function, the function cannot cross 0 from *above* at *both* roots; neither can it cross 0 from *below* at *both* roots.