

CAN CAMP GET YOU INTO A BETTER SECONDARY SCHOOL? A FIELD EXPERIMENT OF TARGETED INSTRUCTION IN KENYA

Beth E. Schueler

(corresponding author)
School of Education and
Human Development
University of Virginia
Charlottesville, VA 22904
beth_schueler@virginia.edu

Daniel Rodríguez-Segura

School of Education and
Human Development
University of Virginia
Charlottesville, VA 22904
dan.rodriguez@virginia.edu

Abstract

Access to quality secondary schooling can be life changing for students in developing contexts. In Kenya, entrance to such schools was historically determined by performance on a high-stakes exam. Understandably then, preparation for this exam is a priority for families and educators. To increase the share of students entering these schools, some educational providers offer targeted instruction for students they believe have a chance of securing a spot. We use a randomized control trial to evaluate the impact of these “symposia” programs—week-long, sleep-away camps where eighth-grade students receive a burst of academic instruction from teachers who are selected based on merit. While similar models have been tested in the United States, less is known about this type of intervention in developing settings. We find these programs were not particularly effective for the average nominated student relative to a typical week of school. However, we find large, positive effects among students attending schools from which few students are nominated for symposia. We provide suggestive evidence that this was because students from low-representation schools had fewer pre-camp practice test resources outside of school. The results have implications for program design and the growing literature on the effectiveness of appropriately targeted individualized instruction.

https://doi.org/10.1162/edfp_a_00322

© 2020 Association for Education Finance and Policy

1. INTRODUCTION

Receiving a secondary school education can be life-changing for young people around the world, and Kenyan students are no exception. Entry into one of the more prestigious and well-resourced Kenyan “national schools” is thought to have particularly important returns in terms of students’ odds of postsecondary and labor market success. As in a number of countries, progression to secondary school and consideration for a seat at a national school is largely determined by performance on a high-stakes exam taken at the end of eighth grade. Students, parents, educators and policy makers therefore have a keen interest in ensuring that students who have a chance at receiving a spot in one of these elite schools are as well prepared for the exam as possible.

One approach to preparing these students involves a relatively short burst of supplementary, intensive, small-group instruction, provided outside of school by high-quality teachers. Similar programs have been studied in the context of the United States, but we provide rare evidence on the efficacy of such a program in the developing world. Specifically, we partnered with a school operator in the Kenyan context called Bridge International Academies (“Bridge”) to conduct a field experimental study of “symposia”—week-long, overnight, academic camps, taught by select teachers—designed to help prepare relatively high-achieving students for the mandatory exam administered to all Kenyan eighth graders as their primary school experience comes to an end.

We find no evidence that these programs had particularly large benefits for the average nominated student relative to attending a typical week of school. However, we find that a subset of our sample made up of students attending schools from which very few students were nominated for the symposia programs (“low-representation schools”) saw substantial gains on the order of 0.30 standard deviation. This finding is not explained by differences in baseline test performance between students from low-representation and higher-representation schools. We find suggestive evidence that the differential effects are due to variation in the level of pre-symposia test preparation resources available to different school communities outside of the school environment. Our findings have implications for educational leaders seeking to effectively target such programs to those with the greatest potential to benefit. We also contribute to the ever-growing literature on the promise of individualized instructional programs.

2. BACKGROUND

A robust literature developed over the past sixty years demonstrates the dramatic returns on a range of important short- and long-run outcomes that result from access to education worldwide (Psacharopoulos and Patrinos 2018). Returns are highest in developing countries, and the context we study—Kenya—is no exception. Kenya has made progress in recent years toward meeting Millenium Development Goals by lowering child mortality and increasing primary school enrollment (World Bank 2019a, b). However, the nation still struggles with a high poverty rate (37 percent of the population lives on less than \$1.90 per day) and low educational attainment at the secondary and tertiary levels (World Bank 2018a). Despite progress on both primary and secondary school enrollments after a 2003 constitutional change making primary education free and compulsory for all, and a 2008 push for free secondary education, net enrollment in secondary school is only 51 percent (World Bank 2018b), and the tertiary school

enrollment rate is a mere 4 percent (World Bank 2018b). Given that Kenya has one of the fastest growing economies in sub-Saharan Africa and that it has a relatively small educated pool, the conditions are right for education to be a high-yield investment in this context. It is therefore not surprising that scholars have previously estimated that secondary school attendance improves Kenyan students' cognitive abilities, employment prospects and wages, and probability of avoiding teen pregnancy (Moraa 2014; Shimada, Khan, and Wakano 2016; Ozier 2018).

However, access to secondary education in this context has largely been determined by student performance on a mandatory high-stakes exam—the Kenya Certificate of Primary Education (KCPE)—taken at the end of primary school. As of 2019, the Kenyan Government changed this policy to allow all students to transition from primary to secondary school. However, historically this exam has determined not only access to, but also type of, secondary school (e.g., elite, provincial, and district schools). Specifically, a certain level of performance has been required to be considered for and guaranteed a spot in one of the country's elite “national” schools. These national schools have better facilities and course offerings than other schools and also offer students access to more highly educated and experienced teachers, as well as a higher-achieving peer group (Lucas and Mbiti 2014). Researchers have estimated that passing the KCPE is associated with an 11 percent increase in wages, while performing well enough to gain entry to and completing secondary education at a national school represents an increase of private returns of 22 percent (Moraa 2014).

Whereas Lucas and Mbiti (2014) find no benefit of admission to these national schools on overall secondary school test scores or graduation, they do find positive effects on Kiswahili test scores. Additionally, Ozier (2018) points out that the Lucas and Mbiti study may not have the power to detect the relatively modest size of effects we might expect on exam scores overall based on previous studies from other contexts. Furthermore, the Lucas and Mbiti analysis is limited to secondary school outcomes, while others find evidence that these schools have returns on postsecondary and labor market outcomes, as described above. After all, national schools are one of the main pathways into university and the highly sought-after and well-remunerated civil service jobs (Glewwe, Kremer, and Moulin 2009). Furthermore, the KCPE also plays a role in admissions for the second tier of Kenyan “provincial” schools, even if students do not score high enough for the national schools.

Therefore, students, parents, educators, and policy makers have a keen interest in learning how best to prepare students who have a decent shot at gaining entry into one of these national schools for the qualifying examination and ultimately success in secondary school. In meta-analyses of education studies in developing contexts, scholars report that supply-side policies have generally proved insufficient when they are not accompanied by other policies that changed the daily schooling experiences of children (Masino and Niño-Zarazúa 2016; Murnane and Ganimian 2017). Muralidharan, Sing, and Ganimian (2019) point out that expansion of education services in developing countries without proper instructional support does not guarantee corresponding levels of skill development and learning, particularly for students from lower-income backgrounds. Consistent with this theme, in the Kenyan context, Lloyd, Mensch, and Clark (2000) found that factors reflecting the school experience are the greatest predictors of student progression in school. Vermeersch and Kremer (2005) find that

subsidized school meals increased test scores for Kenyan students but only in schools with better trained teachers, providing further evidence of the importance of school quality.

One approach to providing the necessary preparation for students is through individualized instructional experiences that are provided by talented teachers and more tailored to student needs than the instruction they receive in a typical school day. The theory is that grouping students based on prior achievement and limiting the number of students under a teacher's charge greatly simplifies the teacher's task and reduces the need for within-class differentiation of instruction. In international contexts, there is some evidence suggesting that tutoring represents a promising approach to individualized instruction. Notably, Banerjee et al. (2007) find that remedial tutoring targeting third and fourth graders who had not yet achieved mastery of their respective grade's material in India showed large gains concentrated among the lowest-performing students. Dang and Rogers (2008) conduct a meta-analysis of the broader literature and argue that among studies controlling for endogeneity and selection, private tutoring has been shown to have a positive relationship with academic performance in India, Israel, Japan, United States, and Vietnam. However, it appears that program design matters, given that not all tutoring programs have demonstrated success internationally (e.g., Cabezas, Cuesta, and Gallego 2011; Jayachandran 2014; Song, Loewenstein, and Shi 2018).

This is consistent with the evidence from the U.S. context, which suggests that low-intensity tutoring (fewer hours, higher student–teacher ratios, less differentiation) has less success (e.g., Heinrich, Meyer, and Whitten 2010; Heinrich et al. 2014) than so called “high-dosage tutoring,” for which a persuasive and robust literature documents substantial benefits (e.g., Cook et al. 2014; Fryer 2014, 2016; Kraft 2015; Chabrier, Cohodes, and Oreopoulos 2016). Tutor quality also likely influences the success of such programs, given the known importance of teacher effectiveness in explaining and improving student academic achievement and long-term success (e.g., Rivkin, Hanushek, and Kain, 2005; Chetty et al. 2011). However, the variation and impact of teacher quality on student outcomes is not as well understood in developing countries as it is in the U.S. context (Glewwe and Muralidharan 2016).

Despite impressive impacts and cost–benefit ratios of high-dosage tutoring (Harris 2009), concerns persist about the scalability of programs with one-to-one or two-to-one student–teacher ratios (Bloom 1984). This has prompted some school systems to adopt intensive small group instructional programs, led by high-quality teachers, for targeted groups of students based on ability. For example, “vacation academy” programs in struggling Massachusetts schools that provide students on the cusp of academic proficiency with instruction from teachers selected on merit in a single subject over week-long vacation breaks have demonstrated positive results in both experimental (Schueler 2018) and quasi-experimental studies (Schueler, Goodman, and Deming 2017). These programs appear more easily scalable than high-dosage tutoring models given that they have somewhat higher student–teacher ratios (i.e., ten-to-one). To our knowledge, this type of program has not yet been evaluated in a developing context, such as Kenya, where it might stand to greatly benefit students but where a variety of contextual factors (e.g., supply and demand for test preparation, test-based

accountability systems, socioeconomic differences between student populations) could moderate the program effectiveness that has been demonstrated in domestic settings.

The related literature from relevant contexts is fairly small. Muralidharan, Singh, and Ganimian (2019) analyze a randomized trial of small-group tutoring combined with computer-aided instruction in India and find large benefits but are not able to disentangle the effects of technology from the effects of small group instruction. Banerjee et al. (2017) study day camps in the context of India but these are remedial programs taught by volunteers and target students in third to fifth grades, rather than programs focused on promoting high school admission among those eighth graders who educators perceive are likely to gain entry with extra support.

Therefore, we study one such intensive instructional program carried out by teachers selected based on teaching merit and designed to target students likely to benefit from such a program, in the context of Kenya. Because these programs have proved especially effective with disadvantaged groups of students in the United States, and given that in this context some students may—regardless of their access to the program we study—have more or less access to resources preparing them to succeed on the KCPE, we also study whether the program was more effective among those who appear to have fewer preparatory resources outside of school in order to learn how the provision of these programs might be most effectively targeted to the students who stand to benefit most.

3. SYMPOSIA

Setting

Our study was conducted in partnership with Bridge, an organization that seeks to expand access to quality schooling for children through the operation of a network of schools, including both government schools operated as part of public–private partnerships and low-cost private schools. Since 2009, this network has served over 500,000 children through hundreds of schools (called “Academies”) in India, Kenya, Liberia, Nigeria, and Uganda (Bridge 2018). The Bridge model has previously been independently evaluated in the Liberian context (Romero, Sandefur, and Sandholtz 2019) and is currently undergoing independent evaluation in Kenya (World Bank 2017). The symposia we study were conducted in the East African nation of Kenya, in the city of Nakuru, located approximately 170 km from the capital city of Nairobi (although students came from Bridge academies across the country).

All Kenyan students in the eighth year of their primary school education are required to take the KCPE, a standardized examination administered by the Kenyan Ministry of Education that is used to certify that a student has completed primary schooling. The KCPE covers five subjects: math, English, Kiswahili, social and religious studies, and science. The exam was high stakes at the time of the study, in that students who did not pass were not able to move on to secondary schooling, and scores were used to place students into different kinds of secondary education (such as vocational training schools or the more prestigious national schools).

Given the stakes associated with the KCPE, Bridge operates overnight academic camps called “symposia” to prepare its students for the exam. The symposia we study

were weeklong programs that occurred during September and October of 2018, one to two months before the 2018 KCPE administration. For these programs, Bridge targeted relatively high-performing students from across their network who they believed, with symposium-based extra preparation, had the potential to earn KCPE scores that would qualify them for entry into one of the country's more prestigious national schools, or at least one of the second tier provincial schools where admission is also determined by KCPE score. In other words, they selected the students they believed would benefit most from the program.

Student Nomination

Bridge nominated 957 students for the symposia from 235 Bridge schools located in various regions across Kenya, representing 75 percent of the total Bridge schools that had students sitting for the KCPE in 2018. To nominate students for the program, Bridge identified those students who were in the top performance quartile of all Bridge students in Kenya, based on Bridge-administered practice KCPE exam scores. These practice exams were taken from previous years of official KCPE administrations. More specifically, Bridge created an index for all its grade 8 pupils based on their average practice KCPE score at baseline, across all five subjects and up to four 2018 practice test administrations, from mid-term and end-term periods. Students missing all four scores for any single subject were excluded from the nomination process. The top 957 students on this index were nominated for symposia participation. This population's performance on practice KCPE exams was similar to the performance of the top quartile of KCPE performers nationwide. On average, students in our sample earned a score of 357 on baseline practice KCPE exams out of a maximum of 500 points.

Importantly, students were nominated on the basis of their baseline performance relative to the full distribution of students across the Bridge network rather than their performance relative to students at their own schools. One result of this decision is that some schools had a greater number and share of their students nominated for the symposia than other schools. In figure A.1, which is available in a separate online appendix that can be accessed on *Education Finance and Policy's* Web site at https://doi.org/10.1162/edfp_a_00322, we display the number of students in our sample on the y -axis by the number of students nominated from a given school on the x -axis. For example, the second bar on the left of figure A.1 indicates there are 112 students in our analytic sample who came from schools where they were one of only two students from that school nominated for the symposia. Schools had between one and nineteen students nominated from their school. The median school had three of their students nominated for symposia. We later describe schools that had three or fewer students nominated as "low-representation schools."

Experimental Design

The 957 nominated students were divided into twelve strata based on their baseline performance. Randomization was done within strata to improve statistical power, leaving 478 students in the control group and 479 in the treatment group. The intra-cluster correlation of total baseline KCPE score within randomization strata was 0.98. After accounting for the inclusion of covariates that explain 60 percent of the variation in the

outcome, the sample size and within-strata randomization provided 0.80 power to detect a minimum program effect of 0.07 standard deviation on test score outcomes. This is similar to the effect of small group math instruction provided at “vacation academies” in low-performing schools in Springfield, Massachusetts (Schueler 2018). In table A.1, available in the online appendix, we describe the treatment and control groups, illustrating that randomization successfully generated two groups that were balanced on all baseline characteristics available, including prior achievement (overall and by subject).

Student Recruitment

Bridge sent each student who was selected for the treatment group a symposium invitation letter and required parents to return a consent form prior to participation. Parents were notified that all aspects associated with attending the program, including transportation to and from the camps, would be provided to students free of charge. Attending the symposia was branded as an honor for the students given their relatively high performance on the practice exams, as well as an opportunity to gain further training that could be helpful for admission into a national school and potentially even the receipt of scholarship opportunities. Despite the potential benefits of the program, we were still surprised to learn that 95 percent of all students selected for the treatment group actually returned consent forms and attended a symposium, and no students from the control group attended, for an overall compliance rate of 97.5 percent. This high compliance rate simplifies our analysis and the interpretation of results.

Programming

In 2018, Bridge ran six weeklong sessions of symposia over the course of a month and a half. Students were sorted into sessions on the basis of their baseline performance. Each randomization stratum was paired with one other adjacent stratum and placed in a single session. Each session served approximately eighty students. Fourteen students were able to switch sessions due to individual situations but we do not have data on which individual students switched. Given that these were supervised overnight programs, students who came to the camps had nearly perfect attendance in the daily classes and activities, except for three students who reportedly got sick during the program.

The symposia treatment consisted of five-day overnight academic camps. The programs were held at a site that had formerly been used as a private boarding high school and is currently used as an institute for training Bridge teachers and school managers. The facilities included classrooms, sleeping areas, playgrounds, and a dining hall. For academic instruction, students were grouped into classes of at most twenty students, and were further grouped into small teams of four or five students for various activities. Each class had its own classroom and pair of teachers, one for humanities and one for science and math. These teachers taught the same set of students throughout the symposia, taking turns acting as the lead and assistant teacher, for an overall student–teacher ratio of ten-to-one. The symposia were taught in English, except for the class sessions covering Kiswahili. The daily schedule included 9 hours and 10 minutes of academic instruction, for a total of 45 hours and 50 minutes of instruction over the

course of the week. On most days of symposium, this was broken up into 50-minute segments for each subject. However, two full days were dedicated to single subjects: science and social studies.

The main instructional method for math and the languages was the distribution of preprepared handouts used to review concepts and vocabulary, and the completion of exercises and essay writing practice. The main instructional method for science and social studies was printed flashcards and pairs or small group activities in a more interactive setting. After lunch, to keep students engaged, students were asked to develop their own presentations in groups without using any reference materials. Assessments were also used to check progress. For all subjects, students returned home with handouts for further practice.

Students also received three meals a day and regular breaks. During breaks and in the evenings, students participated in extracurricular games and activities and watched films meant to provide academic motivation, such as a film about a student spelling bee competition and another film about Ben Carson, the African American politician and former neurosurgeon.

Staffing

A total of eighteen instructors rotated through teaching the various sessions of symposia (eight teachers per session). These instructors also served as Bridge Academy teachers during the regular academic year. The teachers were selected based on Bridge's perception that they were particularly effective educators. Approximately 40 percent had taught at symposia in previous years. Teachers who were new to the symposia in 2018 were paired with a teacher who had previously taught at a symposium. The teachers were instructed that the symposia attendees were high-performing students needing to acquire better tools to prepare for the KCPE. Bridge, following their typical model, provided teachers with a curriculum and trained them on what to teach as well as rules and regulations, although the teachers were given more flexibility than teachers in a typical Bridge Academy classroom during the regular school year.

In addition to the academic teachers, other on-site, nonteaching staff consisted of a program manager and four counselors to help with motivational and enjoyable nonacademic activities during breaks and after learning hours, as well as managing student behavior. There was also a nurse on-site and staff in charge of food preparation and accommodation logistics.

Control Condition

Importantly, control group students attended school as usual while the symposia were occurring. Thus our findings represent the effect of symposia relative to a typical week of school rather than the effect of symposia relative to a lack of structured instructional time. In a typical school day, Bridge students receive 8 hours and 25 minutes of instruction, Monday through Friday, and another 5 hours and 25 minutes on Saturdays. Therefore, over the course of the five-day programs, symposia students actually were offered about 1 hour and 40 minutes less instruction time than the average control group student received. Although we do not have data on the school attendance of the control group children during this period, we do know that the baseline attendance rate

of our analytic sample was 90 percent (both for the treatment and control group). If we assume that the control group attended school at a similar rate during the period when symposia were in session, they would have received about 3 fewer hours of instruction than the treatment group. Therefore, the treatment/control contrast represents only a small difference in learning time. The typical eighth-grade average pupil–teacher ratio for Bridge schools in Kenya is roughly twelve students, and classes are typically led by a single teacher at a time. Therefore, the ten-to-one student–teacher ratio at the symposia was slightly lower than the ratio for the control group students, though not dramatically so. Additionally, a period in a typical Bridge school day is 35 minutes (shorter than the 50-minute periods at the symposia).

Therefore, when thinking about the treatment-control contrast for the week of symposia, the average treatment group student had a teacher selected based on merit (versus a Bridge teacher of average ability), a class of peers that had been grouped based on ability and grade level (rather than grade level alone), instruction tailored to ability level and focused on KCPE content, a teacher who was given some additional instructional autonomy relative to the highly standardized typical Bridge model, a slightly smaller class (ten-to-one vs. twelve-to-one student–teacher ratio), and a sleep-away “camp” experience. They had also been identified and celebrated as a high performer in the process of receiving the symposia invitation.

4. METHODS

Data

We relied on administrative data provided by Bridge for the full analytic sample of 957 students who were nominated for the symposia in 2018. At the individual student level, these data include information on student randomization and symposia attendance, as well as post-symposia KCPE scores overall and by subject. We also have student-level demographic variables including gender, age, years a student attended a Bridge school, as well as pre-symposia, baseline school attendance records (number of days present), pre-symposia baseline scores on in-class evaluations mostly consisting of practice KCPE exams taken in eighth grade overall and by subject, and each student’s placement exam score when first entering Bridge. We also know the school each student attended. We have universal coverage for nearly all of these variables except Bridge placement scores, which we are missing for 7.94 percent of our sample. For that variable, we impute the average placement score for students missing on that variable and include a binary indicator for whether a student was missing. Additionally, by providing estimates from unconstrained models we will show below that our results are not driven by the inclusion of our control variables.

For baseline, practice KCPE scores, everyone in our sample had at least one baseline score in each subject. In other words, we have no missingness on any of our baseline scores. However, some students had multiple baseline scores within subject. Students could have up to four baseline scores per subject, one for each of four testing rounds. Additionally, all students in our sample took all subjects during the same testing round at least once. For our main results, the baseline score variable represents the average of all available baseline scores and we also control for the total number of scores available across all subjects to control for possible systematic differences in the overall

amount of practice test taking. We also check that our results are not due to differential practice exam taking in a few ways. Specifically, we run versions of our models where we: (1) replace the average baseline scores with the earliest score available, (2) control for the number of scores available by subject, and (3) control for the number of scores by practice test administration (test 1, test 2, test 3, test 4). The first addresses potential concerns about retaking, the second addresses potential concerns that fewer scores on some subjects are different than fewer on others, and the third addresses potential concerns that fewer scores at some points in time are different than fewer at other times. None of these changes affects our conclusions. Results from these robustness checks are available on request.

Finally, Bridge also provided school-level data for the schools represented in our analytic sample, including information on each school's urban/peri-urban/rural location, average baseline scores of all eighth graders, total enrollment, gender breakdown, and average attendance records for students, teachers, and managers. Unfortunately, the school-level data have a greater degree of missingness than our student-level data—described in greater detail below—which limits our ability to explore variation in effects on these particular school-level characteristics.

Outcome Measures

Our outcomes focus on KCPE performance at the end of eighth grade, including both continuous and binary measures, indicating whether a student performed at or above a given policy-relevant threshold. No students were missing final KCPE scores. For the continuous measure, we standardized KCPE test scores using the full analytic sample to have a mean of zero and a standard deviation of one. We do this for total KCPE scores as well as scores by subject. We also examined effects on a set of three binary outcomes representing different performance thresholds on the KCPE because these thresholds had particular policy relevance. For these binary outcomes, we generated three variables all focused on total, rather than subject-specific, KCPE scores, and each representing a different performance threshold: (1) earning 250 points or more, which is necessary to pass the KCPE exam, (2) earning 300 points or more, which is the score necessary to be considered for a spot at a national school, and (3) earning 400 points or more, which is the cutoff that has traditionally guaranteed a spot in a national school. We therefore were interested to see whether effects were concentrated at a particular threshold (perhaps driven by accountability pressure) and believed that the results of this analysis would be helpful in quantifying and communicating the real-world effects of the program for children's educational trajectories. These binary outcomes were coded one if a student scored at or above the relevant performance threshold and zero if they scored below that threshold.

Empirical Strategy

We begin by estimating the causal effect of receiving an invitation to attend a symposium, or assignment to the treatment group, on achievement outcomes by generating intent-to-treat (ITT) estimates with the following model:

$$KCPE_i = \beta_0 + \beta_1 Treatment_i + X_i + \alpha_s + \varepsilon_i,$$

where $KCPE_i$ refers to our outcome measures described above for student i . $Treatment_i$ is a dummy indicator for whether the student was invited to the symposia (regardless of whether he or she attended), and X_i refers to a vector of individual-level covariates, including the student's gender, age, years at Bridge, baseline school attendance record, average baseline scores on the Bridge-administered KCPE practice exams, number of baseline practice scores recorded, and placement exam score taken when the student first entered a Bridge school. We display results from both unconditional models and models that include these controls in order to improve the precision of our causal estimates. Finally, we include fixed effects for randomization strata s and cluster standard errors at the strata level in all models. Given the binary nature of the dependent variables representing various thresholds, we also ran probit models for these outcomes. However, because our interpretation of the results was not dependent on this modeling choice, we display results from the linear specifications in the paper for consistency and ease of interpretation.

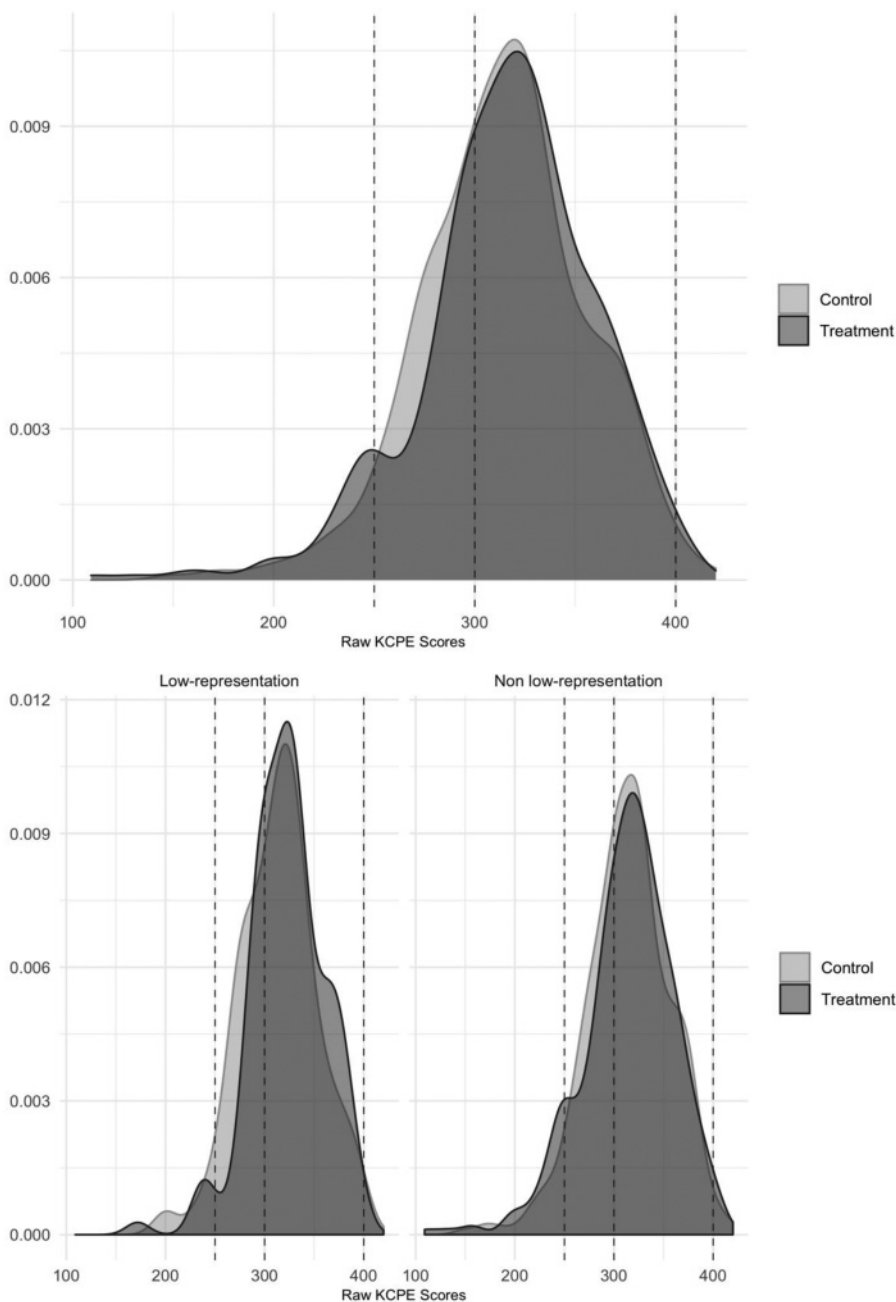
As we discussed earlier, 95 percent of invited students attended the symposia and none of the control group students attended. In table A.2, available in the online appendix, we show that invited noncompliers were less likely to be female but were otherwise similar to compliers. However, because an invitation to the symposia did not guarantee attendance at the programs, we also generated treatment-on-the-treated (TOT) estimates to assess the effect of attending symposia for those who were induced to attend by the symposium invitation. Given that none of the control group students gained access to the treatment, in this case, the local average treatment effect simplifies to the TOT. We calculate these estimates by predicting attendance based on assignment to the treatment group and then regressing predicted attendance on our outcomes. As we illustrate in online Appendix tables A.3 and A.4, our TOT estimates are very similar to the ITT results and do not change our overarching conclusions. Therefore, our central results focus exclusively on the ITT estimates given these are the most conservative estimates and the causal interpretation of these results is most clear.

5. FINDINGS

Symposia Impacts Overall

We start by presenting raw achievement data for the full analytic sample in figure 1, a density plot of standardized KCPE scores by treatment group assignment. This figure shows that the scores for the treatment group (represented in dark gray) are shifted slightly to the right of the control group (represented in light gray). However, the differences are subtle and appear to be concentrated in particular parts of the distribution (around the 250 and 350 thresholds), a finding that we will return to when presenting our formal estimates.

Next, we display the results from our ITT model described above. The top panel of table 1 provides estimates of symposia impacts on standardized KCPE scores overall and by subject, estimated both without and with covariates in our models. Results with coefficients reported for the full set of covariates can be found in online Appendix table A.7. We observe small positive effects on total KCPE scores on the order of 0.05 standard deviation but these effects do not achieve statistical significance. The effect is the same when we estimate it including covariates in our model, and remains statistically



Note: KCPE = Kenya Certificate of Primary Education.

Figure 1. Density Plot of Total Raw Kenya Certificate of Primary Education (KCPE) Scores By Experimental Assignment

no different from zero. Given that we do not have the statistical power to detect effects of such a small magnitude, we cannot be sure whether these effects are true program impacts or due to random chance. Similarly, we find small, positive, nonsignificant effects within all subjects (ranging from 0.05 to 0.08 standard deviation when including

Table 1. The Effect of a Symposium Invitation on Kenya Certificate of Primary Education (KCPE) Scores Overall, By Subject, and By School of Origin Representation (N = 957)

	Total		English		Kiswahili		Math		Science		SSRE	
Treatment	0.05 (0.06)	—	0.09 (0.09)	0.06 (0.08)	0.08 (0.06)	0.08 (0.06)	0.03 (0.06)	0.05 (0.05)	0.05 (0.05)	0.06 (0.05)	−0.04 (0.07)	−0.03 (0.07)
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Treatment × low-representation	0.26** (0.08)	0.25*** (0.07)	0.27** (0.11)	0.24** (0.11)	0.13 (0.11)	0.12 (0.11)	0.27** (0.09)	0.27*** (0.08)	0.11 (0.12)	0.13 (0.13)	0.22** (0.09)	0.24** (0.10)
Low-representation school	0.11 (0.10)	0.09 (0.09)	−0.00 (0.11)	−0.04 (0.11)	0.20** (0.07)	0.20** (0.08)	−0.03 (0.09)	−0.03 (0.09)	0.23* (0.12)	0.18 (0.11)	0.11 (0.10)	0.08 (0.11)
Treatment	−0.01 (0.06)	−0.02 (0.06)	0.02 (0.08)	0.00 (0.08)	0.04 (0.07)	0.05 (0.07)	−0.04 (0.07)	−0.02 (0.06)	0.02 (0.07)	0.03 (0.07)	−0.10 (0.09)	−0.09 (0.08)
Singleton												
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Control mean	−0.022		−0.041		−0.035		−0.012		−0.021		0.025	

Notes: Each column represents a separate regression in which we regress the listed outcome on a binary indicator for treatment group assignment and randomization strata for the top panel. For the bottom panel, we also include a binary indicator for coming from a low-representation school and an interaction between that variable and the treatment indicator. Low-representation school is defined as having three or fewer students in the analysis sample. All models include randomization strata fixed effects. The models with controls include gender, age, years at Bridge, baseline attendance, Bridge placement score, average baseline practice KCPE score of the subject being regressed, and total number of baseline practice KCPE scores. All coefficients are displayed in standard deviation units. Scores have been standardized to the analytic sample. Standard errors are clustered at the randomization strata-level and presented in parentheses. SSRE = Social Studies and Religious Education.

*, **, ***, ****, *****: $p < 0.1$; $p < 0.05$; $p < 0.01$.

Table 2. The Effect of a Symposium Invitation on Kenya Certificate of Primary Education (KCPE) Score Thresholds, Overall and By School of Origin Representation ($N = 957$)

	Over 250		Over 300		Over 400	
Treatment	-0.03*	-0.03**	0.02	0.02	-0.00	-0.00
	(0.02)	(0.01)	(0.03)	(0.03)	(0.00)	(0.01)
Controls	N	Y	N	Y	N	Y
Treatment $\times$ low-representation	0.03	0.03	0.11**	0.12**	-0.01	-0.00
	(0.03)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)
Low-representation school	0.03	0.03	0.00	-0.00	0.00	0.00
	(0.03)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)
Treatment	-0.04**	-0.04**	-0.00	-0.01	-0.00	0.00
	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)
Controls	N	Y	N	Y	N	Y
Control mean	0.952		0.663		0.010	

Notes: Each column in the top panel represents a separate regression in which we regress a binary indicator for whether a student scored higher than the specified threshold on the KCPE on a binary indicator for being assigned to the treatment group and fixed effects for randomization strata. For the bottom panel, we also include a binary indicator for coming from a low-representation school and an interaction between that variable and the treatment indicator. Low-representation school is defined as having three or fewer students in the analysis sample. All models include randomization strata fixed effects. The models with controls include gender, age, years at Bridge, baseline attendance, Bridge placement score, average baseline practice overall KCPE score, and total number of baseline practice KCPE scores. The coefficients can be interpreted as the change in probability of scoring above a given threshold in percentage point units. Standard errors are clustered at the randomization strata-level and presented in parentheses.

* $p < 0.1$; ** $p < 0.05$.

controls) except for social studies/religious education, where we observe small, negative, nonsignificant effects.

In the top panel of table 2, we provide results for the binary threshold outcomes that are consistent across versions of our models excluding and including covariates. Results with coefficients reported for the full set of covariates can be found in online Appendix table A.8. We find that symposia had no impact on a pupil's likelihood of scoring 400 or higher (guaranteeing entry into a national school) and only a small, nonsignificant, positive effect on the likelihood of scoring over 300 or higher (guaranteeing consideration for a national school). There is limited variation in the "over 400" outcome given that very few students achieved scores above this threshold overall, making it less likely that we have power to detect small effects. We also find a small negative effect that does achieve formal statistical significance on the likelihood of scoring 250 or higher (passing the KCPE). Visual evidence of this result can be found in figure 1, which shows a bump in the treatment group distribution right below the 250 cutoff. We consider likely explanations for this in the discussion section below.

Impacts for Students from Low-Representation Schools

As we have already described, there was variation in the number of students that were nominated from each of the 235 Bridge schools that are represented in our analytic sample. To examine whether the effects of symposia programming varied depending on how many students came from a sending school, we generated a binary variable equal to one if a student came from a school that had three or fewer students nominated for symposia (three being the school-level median number of students nominated from a given school in our sample). We describe these students as coming from "low-representation

Table 3. Balance Test of School-Level Characteristics Between Low- and Non-Low-Representation Schools ($N = 235$)

	Low-representation ($N = 134$)			Non low-representation ($N = 101$)			Low — Non-low-representation	
	Mean	SD	N	Mean	SD	N		p -value
Total enrollment	146.49	(64.76)	122	207.03	(79.91)	98	−60.54***	0.00
Urban	0.25	(0.44)	55	0.47	(0.50)	57	−0.22**	0.02
Peri-urban	0.11	(0.31)	55	0.07	(0.26)	57	0.04	0.47
Rural	0.64	(0.49)	55	0.46	(0.50)	57	0.018*	0.06
Share of school invited to symposia	0.01	(0.01)	122	0.04	(0.02)	98	−0.02***	0.00
N of nominated students from school	1.84	(0.75)	134	7.04	(3.45)	101	−5.20***	0.00
Share female	0.48	(0.05)	122	0.48	(0.04)	98	−0.00	0.81
Baseline total KCPE score	296.63	(31.59)	81	318.16	(32.52)	78	−21.52***	0.00
Invitees-school baseline difference	47.99	(32.12)	81	40.61	(28.57)	78	7.39	0.13
Years of school being Bridge	5.79	(1.18)	81	6.14	(1.25)	78	−0.35*	0.07
Principal attendance record	0.46	(0.34)	122	0.48	(0.37)	98	−0.02	0.66
Teacher attendance record	0.70	(0.07)	122	0.70	(0.07)	98	−0.01	0.45
Student attendance record	0.43	(0.13)	122	0.45	(0.12)	98	−0.02	0.21
N of school-level variables missing	3.01	(3.06)	134	1.91	(2.49)	101	1.10***	0.00

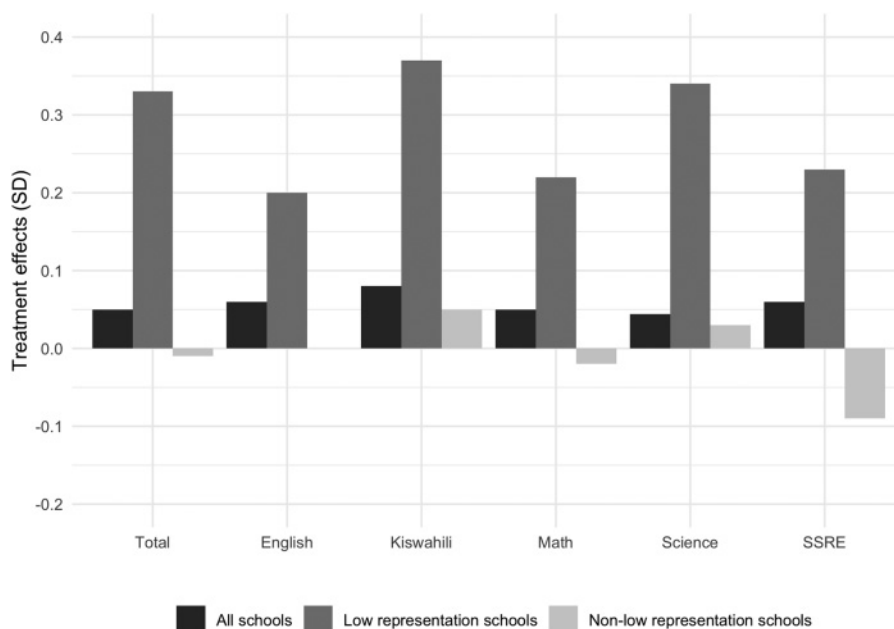
Notes: Here sample sizes represent the number of schools rather than number of students. Low-representation school is defined as having three or fewer students in the analysis sample. Difference in baseline represents the difference by school in the average baseline score of symposia nominees and the rest of the eighth graders in that school. SD = standard deviation; KCPE = Kenya Certificate of Primary Education.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

schools” given the schools sending a greater number of students had greater representation at the symposia. In table 3, we show that these low-representation schools not only had fewer students nominated in terms of absolute numbers but also a smaller share of their overall schoolwide populations.

In figure 1, we provide density plots of raw final KCPE scores by treatment and control assignment separately for students from low-representation and non-low-representation schools. Already this descriptive evidence suggests that symposia gains are concentrated among students from these low-representation schools. When we conduct a Kolmogorov-Smirnov test of distributional equivalence, we find no differences between the treatment and control group among students from non-low-representation schools ($p = 0.79$). However, for students from low-representation schools, the treatment-control difference is larger in magnitude and approaches statistical significance, despite our limited statistical power with this subgroup ($p = 0.19$).

Next, we interact the binary “low-representation” indicator with our treatment dummy and include the interaction and both the main effect of treatment and coming from a low-representation school in the main model described above. We display results in the bottom panel of table 1, showing that the effects of symposia were substantially and significantly larger for students from the low-representation schools of origin than students sent from schools with a greater number of nominated students. This is true for total KCPE scores but the differences are particularly concentrated in English, math, and social studies and religious education (the interaction terms for Kiswahili and science are sizable but do not achieve statistical significance). In figure 2, we display the total effects of an invitation to the symposia for the full sample, as well as the subsamples of students from low-representation schools and non-low-representation schools.

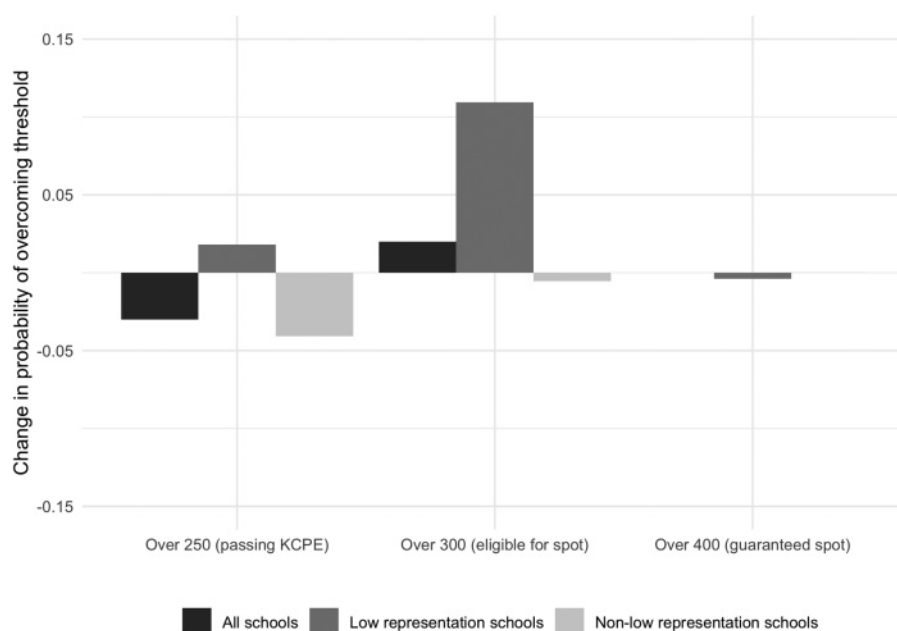


Notes: The bars represent the full effect of treatment by group. The bar for low representation schools is the sum of the coefficient on treatment, on low representation, and on the interaction between these two. The bar for non-low representation is only the coefficient on treatment on the regression that included the interaction between low representation and treatment. The bar for all schools represents the coefficient for the regression without any interactions. KCPE scores are in standard deviation units. SSRE = Social Studies and Religious Education.

Figure 2. The Effect of Symposium Invitation on Kenya Certificate of Primary Education (KCPE) Scores, By Subject and School of Origin Representation

We do this overall and by subject. Symposia had large effects on overall KCPE scores for students from low-representation schools on the order of 0.33 standard deviation (the total of the three relevant coefficients), compared with a statistically insignificant -0.01 standard deviation for the rest of the students. As we show in table 2, these findings are not driven by the inclusion of covariates in our models, as the estimates are quite consistent across specifications excluding and including controls. Furthermore, in Appendix figure A.3 (available in the online appendix), we show that these findings are robust to various definitions of low-representation schools, both overall and by subject. The consistency across subjects also provides further confidence that this finding is not simply due to chance.

One possibility is that the larger effects for students from low-representation schools are driven by a phenomenon whereby a student benefits more from the program because of that student's own self-perception based on being the sole person from his or her school to receive a symposium invitation. In other words, perhaps students who are sent on their own either feel special or feel greater pressure to try hard and get as much as possible out of the symposia. We test this possibility by adding a control variable to our model for whether a student was a "singleton" (the only student from his or her school to be invited to a symposium) based on the idea that this phenomenon would be most salient for students who are the sole invitee. In the third column of table 1, we show that the symposia impact is still 0.17 standard deviation larger for



Notes: The bars represent the full effect of treatment by group. The bar for low representation schools is the sum of the coefficient on treatment, on low representation, and on the interaction between these two. The bar for non-low representation is only the coefficient on treatment on the regression that included the interaction between low representation and treatment. The bar for all schools represents the coefficient for the regression without interactions.

Figure 3. The Effect of Symposium Invitation on Kenya Certificate of Primary Education (KCPE) Score Thresholds, By Subject and School of Origin Representation

students from low-representation schools, even after controlling for singleton status. Therefore, the differential effects do not appear to be exclusively the result of some of these students being the sole student from their school to be invited to a symposium. This finding begins to suggest that the differential effects are due to some other school-level rather than student-level phenomenon (something we explore in greater depth below).

We also examine whether symposia effects vary by school of origin representation on the three threshold outcomes of interest, and display the results in the bottom panel of table 2. We find that the symposia did indeed have a larger effect on the likelihood of scoring over 300 (guaranteeing consideration for a national school) for students from low-representation schools than for schools nominating a greater number of students. We find no differences in the effect of symposia on either the 250- or 400-point threshold between these two groups. We display more easily interpretable impacts overall and by subgroup on each threshold outcome in figure 3. There we show that symposia increased the probability that students from low-representation schools scored a 300 or higher by 11 percentage points relative to a control group rate of 66 percent scoring 300 or above. Interestingly, we also find that the small average negative effect of symposia on passing the KCPE reported earlier seems to be fully driven by students from non-low-representation schools (although these differences are not statistically significant). Again, results are consistent between models with and without controls.

Understanding the Differential Effects

Given the sharply different results for students from the low- versus higher-representation schools, one must wonder what makes symposia so much more effective for students from low-representation schools. Although we are not able to definitively pin down the reason(s) for these differences with the available data, we provide suggestive evidence that this difference is due to variation in access to pre-symposia resources for preparing for the KCPE exam outside of school. Specifically, it appears those from low-representation schools had less access to pre-symposia KCPE preparation resources than students from schools that sent a greater number of students, and therefore benefitted more from symposium-based KCPE preparation. It is unlikely that there were dramatic differences in school-based resources given the standardization of Bridge's model across schools and regions. Furthermore, anecdotally, our research partners at Bridge indicated that communities surrounding some Bridge schools have greater access to old KCPE exams from previous years and greater circulation of these exams outside of school than other communities. This results in some students having more opportunities to work on practice exams at home, such that by the time they take the in-school examinations, they may have seen the test or some test questions before. In other communities, parents are less likely to circulate these practice exams outside of school, making in-home test preparation a less common practice.

If the low-representation schools reside in communities where this at-home KCPE preparation is less common, we would expect baseline grades for higher-representation schools to be a noisier proxy for mastery of KCPE material than for low-representation schools. This indeed appears to be the case. In online Appendix figure A.2, we show that all students in our sample, on average, had higher absolute scores on the baseline practice exams than on the actual post-symposia KCPE. However, the difference between average baseline and final KCPE performance was smaller among students from the low-representation schools than from the rest of the schools. More specifically, students from low-representation schools scored 350 points on the baseline exams and 319 points on the final exams, while students from the higher-representation schools scored 359 points on the baseline exams and 314 on the final exams, on average.

To explore this issue further, we calculated a "disappointment index" by subtracting each student's final KCPE score from his or her baseline KCPE score. In other words, a larger value on the disappointment index would suggest that a student's final KCPE exam score was lower than she might have expected based on her practice KCPE. For this exercise, we limit our sample to the control group only in order to avoid picking up on the confounding effect of treatment. In online Appendix figure A.4, we show that the average disappointment index (or difference between baseline and final KCPE) is significantly smaller for students from the low-representation schools than the higher-representation ones. This result is robust to various definitions of low-representation schools (expanding and contracting the threshold window by two students). This finding is consistent with the story that students from the higher-representation schools had baseline scores that were inflated by pre-symposia KCPE preparation and were therefore a less accurate measure of true mastery of the skills assessed by the final KCPE. Students from the low-representation schools, on the other hand, had fewer pre-symposia KCPE preparation resources, which was reflected in their baseline scores and therefore benefitted more from the provision of KCPE preparation at the symposia.

It is important to keep in mind that students were not randomly assigned to low- versus higher-representation schools. Therefore, there could be differences between these schools other than access to test preparation resources that explain the differential symposia impacts. Unfortunately, we do not have complete data on the schoolwide characteristics of all schools in our sample. As we show in table 3, the school covariates display some missingness, unlike the fully populated individual-level data. Within the 235 schools, there are, on average, 2.5 missing values among all twelve school-level data covariates, and we find that low-representation schools have, on average, 1.1 more missing values than the rest of the schools. With the caveat that the data are incomplete and that we have a somewhat higher rate of missingness for low-representation schools than higher-representation ones, we compare the average school-level characteristics of low- versus non-low-representation schools in table 3. We find no differences on common indicators of school quality in developing contexts, including student, teacher, and academy manager attendance rates. This, combined with Bridge's highly standardized delivery model, suggests it is unlikely that major differences in quality across low- and higher-representation schools explain the differential effects across these school types.

However, we do find evidence that these schools differ on other important dimensions. Specifically, on average, the low-representation schools are smaller (by 61 students), more likely to be in a rural location (by 18 percentage points), and lower scoring at baseline (by 21.53 points). Finally, students at these schools were, on average, newer to Bridge than students at the higher-representation schools (by 0.35 year). However, we find no evidence that any of these observable differences are driving our result regarding the large impacts for students from low-representation schools. In online table A.7, we show results from tests of these alternative explanations after we interact each of these school level characteristics with our treatment indicator. We find that within school type (low- and higher-representation schools) students are no more or less likely to benefit from symposia if they come from a rural, higher-achieving, or smaller school.

Furthermore, we do not find that the differential effects are driven by any of the student-level characteristics we observe in our data. We first contrast the characteristics of students at the two types of school in online table A.6 (an exercise for which we have better data coverage than for the school-level analysis described in the previous paragraph). At the student level, among those nominated for symposia, students from the low-representation schools had been at Bridge schools for slightly longer (by 0.22 year or ~80 days), had lower placement scores when they originally entered a Bridge school (by less than a point on a scale of 0 to 27) and, as we have already mentioned, were lower achieving on their baseline KCPE scores (by 8.91 points) than those at the higher-representation schools.

In table 4, we show results from formal tests of interactions between treatment and each of the major student-level demographic and baseline performance characteristics. We do not find the effects vary based on student sex, age, years at Bridge, placement score, pre-symposia attendance, or baseline KCPE score. The lack of significant interactions based on any measure of baseline performance is particularly notable since it eliminates the possibility that students from low-representation schools benefit more from symposia simply because they are lower-achieving students. We also control for all baseline achievement measures in all of our previous models, providing further evidence on this point. Although we do not include results here for the sake of

Table 4. The Effect of Symposium Invitation on Kenya Certificate of Primary Education (KCPE) Scores By Student Characteristics ($N = 957$)

	Female	Age	Years at Bridge	Placement Score	Baseline Attendance	Baseline Total Scores
Treatment $\times$ characteristic	0.05 (0.13)	−0.05 (0.05)	0.06 (0.05)	0.00 (0.02)	−0.35 (0.35)	0.00 (0.00)
Treatment	0.03 (0.07)	0.77 (0.72)	−0.14 (0.20)	−0.01 (0.24)	0.36 (0.33)	−0.20 (0.80)
Characteristic	−0.19* (0.09)	−0.02 (0.03)	0.05 (0.04)	−0.00 (0.01)	0.51 (0.33)	−0.00 (0.00)

Notes: Each column represents a separate model in which we regress total standardized KCPE scores on an interaction between a binary indicator for being assigned to the treatment group and the characteristic listed at the top of the column, a binary indicator for treatment group assignment, the listed characteristic, and fixed effects for randomization strata. All models include randomization strata fixed effects and controls for gender, age, years at Bridge, baseline attendance, Bridge placement score, average baseline practice overall KCPE score, and total number of baseline practice KCPE scores. All coefficients are displayed in standard deviation units. Scores have been standardized to the analytic sample. Standard errors are clustered at the randomization strata-level and presented in parentheses.

parsimony, we have also calculated an Oaxaca-Blinder decomposition and found additional evidence that our results were driven primarily by differential treatment effects for the two groups rather than differences in values of the independent variables (e.g., baseline performance) within the treatment and control groups. This fact pattern lends further credibility to the idea that the differential effects for students from low-representation schools are due to a school- rather than student-level phenomenon.

6. DISCUSSION

We used a field experiment to test the provision of weeklong overnight academic camps—“symposia”—designed to prepare relatively high-achieving Kenyan students for the high-stakes KCPE exams that play a major role in determining students’ secondary school experiences. Our study does not reveal strong evidence that these programs were beneficial for the average invitee relative to a typical week of school. Specifically, we find symposia have a small positive effect on average students’ KCPE performance, but these effects do not achieve formal statistical significance and we do not have the power to detect effects this small. Therefore, we are not able to rule out the possibility that we have observed this result due to chance alone. Perhaps surprisingly, we find, for the average student, symposia had a small negative effect on KCPE passing rates.

However, in order to interpret this result, it is important to keep the counterfactual in mind. Because the control group attended school as usual, this finding simply suggests that students near the passing threshold would have been better off simply attending school rather than symposia. This finding is consistent with the fact that symposia were geared toward higher-achieving students and therefore were probably less likely to focus on basic skills for passing than were teachers in a typical Bridge classroom. It is unlikely this result is driven by discouragement effects among low performers because we find no interaction between baseline achievement and treatment, and because students were randomized into treatment within relatively granular baseline performance bins (and therefore attended symposia with similarly achieving peers). It is possible that our results overstate the program’s effect if high-performing

students who were randomly assigned to the control group became discouraged by not receiving an invitation, though we do not consider this to be a major concern given our results for the average student are not significantly positive. Collectively, these findings suggest that educational leaders should look elsewhere or rework the program design in order to provide effective supplemental support for average high performers on their high-stakes exams.

Although we do not find strong evidence the symposia were particularly beneficial for the average nominated student, we do find these symposia had large positive effects for a subset of students who attend schools from which only a small number of students were nominated for the program. Additionally, we find no negative effects on KCPE pass rates for these students. The magnitude of effects on KCPE performance for this group is substantively meaningful. Average KCPE scores for students from low-representation schools increased from 314 to 328. Brudevold-Newman (2017) found that an increase of this kind on the KCPE exam is associated with a statistically significant approximately 3.4 percentage point increase in the probability of sitting for the secondary school exit exam, which is a strong proxy for completion of secondary school.

Furthermore, we find suggestive evidence that students from low-representation schools receive greater benefits from symposia because they have had fewer opportunities for pre-symposia KCPE preparation outside of school. Although we cannot definitively pinpoint the mechanism, our evidence still provides guidance for educational leaders and policy makers seeking to effectively target programs to those students with the greatest potential to benefit. Leaders may also want to consider new approaches to equalizing the distribution and circulation of test preparation materials and resources to ensure that students from schools not currently well-represented at the symposia can improve their chances of performing well on the KCPE regardless of whether they attend the symposia. Our study adds to a growing literature about the benefits of intensive individualized instructional experiences, delivered by high-quality teachers for targeted student populations that has, up until now, focused more heavily on the United States rather than developing nations.

One limitation of our study is that we did not have the foresight to register a pre-analysis plan. Olken (2015) finds the benefit of conducting pre-identified analysis plans is not always as beneficial as scholars may expect, however, such plans can help limit the risks of reporting spurious results. To minimize these risks without having registered, we ultimately report the full extent of analyses we conducted including several null results. While we do not report every single regression, we report results from all substantive research questions we tested using a specification that represents a finding that is robust to any minor modeling variations we tested in our analysis phase. We also reflected on the extent to which our analyses represented tests we would have included in a pre-analysis plan had we submitted one. Our best guess is that all of our analyses would have been included in such a plan with one important exception: our low-representation analysis. This analysis came about as our findings began to indicate that variation in program effectiveness was better explained by school- rather than student-level characteristics. The chances are small that the low-representation findings are spurious, given that they are large in magnitude, replicate across a variety of academic subjects, are theoretically grounded, resonated with our partners

familiar with the context on the ground, and vary as expected under different definitions of “low-representation.” That said, replication would be valuable for providing further confidence in these particular results.

To the extent that our findings would replicate, based on a back-of-the-envelope calculation for the low-representation group, symposia have a comparable cost-effectiveness to other educational interventions in developing contexts, such as personalized technology-aided instruction. Specifically, symposia appear slightly more cost effective at raising language arts performance and slightly less cost-effective at raising math performance (Muralidharan et al. 2019). However, symposia involved only a week of programming and were therefore especially time-effective. That said, there are other educational interventions with even smaller cost-benefit ratios (McEwan 2015) and it is not impossible that merely distributing preparatory materials to students at low-representation schools could prove to be a more cost-effective alternative. To reiterate, these programs do not appear to be a particularly good use of resources for the average high-scoring Bridge student. Positive cost-benefit ratios appear to depend on effective targeting within the population of high-scoring Bridge students.

There are a number of questions that our study leaves unanswered. For example, we are unable to determine the mechanisms that make symposia effective for students from low-representation schools. In previous studies in the Kenyan context, Duflo, Dupas, and Kremer have found benefits of class size reduction in some cases (2015) and tracking, or the placement of students in a class with other students of similar ability level (2011). It could be lower teacher-student ratios, teachers selected on the basis of perceived merit, ability grouping, teacher autonomy, or a combination of any or all of these program features, are essential to producing the gains we observe for this subgroup.

Future research should examine the effects of such programs on non-test and longer-term outcomes of substantive interest. Given that we are limited to examining effects on KCPE outcomes, we cannot entirely rule out the possibility that the program improves test scores without generating genuine learning gains. If this is the case, it is not impossible that the program simply helps participating students get into better secondary schools while preventing other equally capable students from attending those schools (given that there are a limited number of seats in these schools). However, even if the program provides no true long-term learning benefits for participants, it could still have social value by providing KCPE preparation for those students who receive less of it in their typical schooling environment, thereby equalizing opportunities to receive a high-quality, secondary school experience. This further highlights the importance of appropriately targeting the program to those students with the most limited access to exam preparation.

Furthermore, we study the effectiveness of a program serving relatively high performers. Even the students from the low-representation schools were still in the top quartile of Bridge student performance at baseline. Our findings may not generalize to lower-achieving students. On the other hand, the gains we do find are concentrated among students from schools with a greater share of low performers. The program could have even larger impacts for lower-achieving students, and indeed, in the U.S. context similar programs have demonstrated substantial benefits for students at lower ends of their respective performance distributions. Regardless of these limitations, our

study documents a relatively cost- and time-effective approach that, when effectively targeted, can meaningfully increase the odds that Kenyan students get to secondary school.

ACKNOWLEDGMENTS

The authors thank Sean Geraghty, Tim Sullivan, Hassan Said, Dorcas Kahaki, and Carol Yu of Bridge International Academies; Andrew Brudevold-Newman and Owen Ozier of the World Bank; as well as Daphna Bassok, Ben Castleman, Isaac Mbiti, and seminar participants at the University of Virginia; Art Goldsmith of Washington and Lee University; John Robst of University of South Florida; and our editor and reviewers for their valuable contributions to this research. We are also grateful to the students, families, teachers, and administrators who made this work possible.

REFERENCES

- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives* 31(4): 73–102. <http://dx.doi.org/10.1257/jep.31.4.73>
- Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics* 122(3): 1235–1264. <http://dx.doi.org/10.1162/qjec.122.3.1235>
- Bloom, Benjamin. 1984. The 2 sigma problem. *Educational Researcher* 13(6): 4–16. <http://dx.doi.org/10.3102/0013189X013006004>
- Bridge. 2018. *About us*. Available <https://www.bridgeinternationalacademies.com/who-we-are/>. Accessed 18 May 2021.
- Brudevold-Newman, A. 2017. The impacts of free secondary education: Evidence from Kenya. Unpublished paper, World Bank.
- Cabezas, Veronica, Jose Cuesta, and Francisco Gallego. 2011. Effects of short-term tutoring on cognitive and non-cognitive skills. Santiago, Chile: Abdul Latif Jameel Poverty Action Lab.
- Chabrier, Julia, Sarah Cohodes, and Phillip Oreopoulos. 2016. What can we learn from charter school lotteries? *Journal of Economic Perspectives* 30(3): 57–84. <http://dx.doi.org/10.1257/jep.30.3.57>
- Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics* 126(4): 1593–1660. <http://dx.doi.org/10.1093/qje/qjr041>
- Cook, Phillip, Kenneth Dodge, George Farkas, Roland Fryer, Jonathan Guryan, Jens Ludwig, Susan Mayer, Harold Pollack, and Laurence Steinberg. 2014. The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth. NBER Working Paper No. 19862.
- Dang, Hai-Anh, and Halsey Rogers. 2008. The growing phenomenon of private tutoring. *World Bank Research Observer* 23(2): 161–200. <http://dx.doi.org/10.1093/wbro/lkn004>
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. Peer effects, teacher incentives, and the impact of tracking. *American Economic Review* 101:1739–1774. <http://dx.doi.org/10.1257/aer.101.5.1739>

- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. School governance teacher incentives, and pupil-teacher ratios. *Journal of Public Economics* 123:92–110. <http://dx.doi.org/10.1016/j.jpubeco.2014.11.008>
- Fryer, Roland. 2014. Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics* 129:1355–1407. <http://dx.doi.org/10.1093/qje/qju011>
- Fryer, Roland. 2016. The production of human capital in developed countries: Evidence from 196 randomized field experiments. NBER Working Paper No. 22130.
- 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. <http://dx.doi.org/10.1257/app.1.1.112>
- Glewwe, Paul, and Karthik Muralidharan. 2016. Improving school education outcomes in developing countries. In *Handbook of the economics of education*, edited by Eric Hanushek, Stephen Machin, and Ludger Woessmann, pp. 653–744. Amsterdam: Elsevier.
- Harris, Douglas. 2009. Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis* 31(1): 3–29. <http://dx.doi.org/10.3102/0162373708327524>
- Heinrich, Carolyn, Robert Meyer, and Greg Whitten. 2010. Supplemental education services under No Child Left Behind. *Educational Evaluation and Policy Analysis* 32(2): 273–298. <http://dx.doi.org/10.3102/0162373710361640>
- Heinrich, Carolyn, Patricia Burch, Annalee Good, Rudy Acosta, Huiping Cheng, Marcus Dillender, Christi Kirschbaum, Hiren Nisar, and Mary Stewart. 2014. Improving the implementation and effectiveness of out-of-school-time tutoring. *Journal of Policy Analysis and Management* 33(2): 471–494. <http://dx.doi.org/10.1002/pam.21745>
- Jayachandran, Seema. 2014. Incentives to teach badly: After-school tutoring in developing countries. *Journal of Development Economics* 103: 190–205. <http://dx.doi.org/10.1016/j.jdevco.2014.02.008>
- Kraft, Matthew. 2015. How to make additional time matter: Extending the school day for individualized tutorials. *Education Finance and Policy* 10(1): 81–116. <http://dx.doi.org/10.1016/j.edfp.2015.00152>
- Lloyd, Cynthia, Barbara Mensch, and Wesley Clark. 2000. The effects of primary school quality on school dropout among Kenyan girls and boys. *Comparative Education Review* 44(2): 113–147. <http://dx.doi.org/10.1086/447600>
- Lucas, Adrienne, and Isaac Mbiti. 2014. Effects of school quality on student achievement: Discontinuity evidence from Kenya. *American Economic Journal: Applied Economics* 6(3): 234–263.
- Masino, Serena, and Miguel Niño-Zarazúa. 2016. What works to improve the quality of student learning in developing countries? *International Journal of Educational Development* 48:53–65. <http://dx.doi.org/10.1016/j.ijedudev.2015.11.012>
- McEwan, Patrick. 2015. Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research* 85(3): 353–394. <http://dx.doi.org/10.3102/0034654314553127>
- Moraa, Peter Janet. 2014. The private returns to education in Kenya. Master's dissertation, University of Nairobi, Nairobi, Kenya.

- Muralidharan, Karthik, Abhijeet Singh, and Alejandro Ganimian. 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review* 109(4): 1426–1460. <http://dx.doi.org/10.1257/aer.20171112>
- Murnane, Richard, and Alejandro Ganimian. 2017. Improving educational outcomes in developing countries: Lessons from rigorous evaluations. NBER Working Paper No. 20284.
- Olken, Benjamin. 2015. Promises and perils of pre-analysis plans. *Journal of Economic Perspectives* 29(3): 61–80. <http://dx.doi.org/10.1257/jep.29.3.61>
- Ozier, Owen. 2018. The impact of secondary schooling in Kenya: A regression discontinuity analysis. *Journal of Human Resources* 53(1): 157–188. <http://dx.doi.org/10.3368/jhr.53.1.0915-7407R>
- Psacharopoulos, George, and Harry Patrinos. 2018. Returns to investment in education: A decennial review of the global literature. *Education Economics* 26(5): 445–458. <http://dx.doi.org/10.1080/09645292.2018.1484426>
- Rivkin, Steven, Eric Hanushek, and John Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73(2): 417–458. <http://dx.doi.org/10.1111/j.1468-0262.2005.00584.x>
- Romero, Mauricio, Justin Sandefur, and Wayne Sandholtz. 2019. Outsourcing education: Experimental evidence from Liberia. *American Economic Review* 110(2): 364–400. <http://dx.doi.org/10.1257/aer.20181478>
- Schueler, Beth. 2018. Making the most of school vacation: A field experiment of small group math instruction. *Education Finance and Policy* 15(2): 310–331.
- Schueler, Beth, Joshua Goodman, and David Deming. 2017. Can states take over and turn around school districts? *Educational Evaluation and Policy Analysis* 39(2): 311–332. <http://dx.doi.org/10.3102/0162373716685824>
- Shimada, Kentaro, Zeba Khan, and Ayako Wakano. 2016. An update of the returns to education in Kenya. Osaka School of International Public Policy Discussion Paper 16-18.
- Song, Yang, George Loewenstein, and Yaojiang Shi. 2018. Heterogeneous effects of peer tutoring: Evidence from rural Chinese middle schools. *Research in Economics* 72(1): 33–48. <http://dx.doi.org/10.1016/j.rie.2017.05.002>
- Vermeersch, Christel, and Michael Kremer. 2005. School meals, educational achievement and school competition: Evidence from a randomized evaluation. Washington, DC: World Bank Policy Research Working Paper No. 3523.
- World Bank. 2017. *Kenya: Evaluation of Bridge schools*. Available worldbank.org/en/programs/sief-trust-fund/brief/kenya-early-childhood-education-project. Accessed 18 May 2021.
- World Bank. 2018a. *Decline of global extreme poverty continues but has slowed: World Bank*. Available <https://www.worldbank.org/en/news/press-release/2018/09/19/decline-of-global-extreme-poverty-continues-but-has-slowed-world-bank>. Accessed 18 May 2021.
- World Bank. 2018b. *World development report 2018: Learning to realize education's promise*. Available <https://elibrary.worldbank.org/doi/abs/10.1596/978-1-4648-1096-1>. Accessed 18 May 2021.
- World Bank. 2019a. *School enrollment, primary (% net)*. Available <https://data.worldbank.org/indicator/SE.PRM.NENR>. Accessed 18 May 2021.
- World Bank. 2019b. *Mortality rate, infant (per 1,000 live births)*. Available <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>. Accessed 18 May 2021.