

Evaluation of Bridge International Academies in Kenya*

Stage One Analysis and Plan for Subsequent Analysis

Guthrie Gray-Lobe[†] Anthony Keats[‡] Michael Kremer[§]

Isaac Mbiti[¶] Owen Ozier^{||}

PLEASE DO NOT CIRCULATE

February 1, 2020

*This study was conducted with permissions from the National Commission for Science, Technology, and Innovation (NACOSTI). Additional approvals from relevant County commissioners/County Education officers were also obtained. IRB approval was received from Maseno University Ethics Review Committee (MUERC) and IPA (protocol 7410). The evaluation was supported by the Omidyar Network, the World Bank Strategic Impact Evaluation Fund (SIEF), JPAL Post-Primary Education Initiative, and the Bill and Melinda Gates Foundation. The scholarship program was funded by supported by the Douglas B. Marshall, Jr. Family Foundation. We are grateful to Isaiah Andrews, Joshua Dean, Andrew Ho, Alaka Holla, Seema Jayachandran, Ben Piper, Mauricio Romero, and Justin Sandefur for helpful comments. We are grateful to Maria Bucciarelli, William Blackmon, Celeste Carano, Peter Hawes, Peter Hickman, Naomi Kimani, Jessica Mahoney, and Ben Wekesa for excellent research assistance.

[†]Harvard University, email: graylobe@g.harvard.edu

[‡]Wesleyan University, email: akeats@wesleyan.edu

[§]Harvard University, email: mkremer@fas.harvard.edu

[¶]University of Virginia, email: imbiti@virginia.edu

^{||}World Bank, email: oozier@worldbank.org

Contents

- 1 Introduction 5**

- 2 Description of Bridge International Academies 6**

- 3 The Scholarship Program 8**
 - 3.1 Overview 8
 - 3.2 Scholarship application recruitment and procedures 9
 - 3.3 Randomization 11
 - 3.3.1 Defining sub-groups 11
 - 3.3.2 Stratification 12
 - 3.3.3 Final randomization sample sizes 13
 - 3.4 Endline Sample 13
 - 3.5 Siblings sample 16

- 4 Data 17**
 - 4.1 Pupil Level Data 17
 - 4.1.1 Baseline Data 17
 - 4.1.2 Midline Phone Call Data 19
 - 4.1.3 Endline Data 20
 - 4.1.4 2019 Phone call follow-up 22
 - 4.2 School Level Data 23
 - 4.2.1 Local Education Environment Survey 23

- 5 Comparison between Bridge and Nearby Schools 24**

- 6 Endline Attrition and Covariate Balance 31**

7	Stage One Analysis	40
7.1	The effects of the scholarship on educational environment	41
7.2	The effect on <i>paying for</i> Bridge	46
7.3	Sibling spillover effects on attendance	46
7.4	Effect of scholarship extension on enrollment	52
7.5	Effects of treatment on grade progression	57
8	Intent of Plan for Subsequent Analysis	64
8.1	Research team data access	65
9	Empirical Framework for Estimation of Causal Effects of the Scholarship	67
9.1	The Scholarship Effect	68
9.2	Decomposing the Scholarship LATE	71
9.2.1	Decomposing the Scholarship LATE	73
9.2.2	Statistical Power for Decomposition	75
9.2.3	Plug-in Estimators	76
9.3	Income effect of the scholarship	81
9.3.1	Plug-in Estimators for direct effect of the scholarship	85
9.4	Interpretation of effect and alternative specifications	86
9.5	Controls	88
10	Outcomes	89
10.1	Academic subject test scores	89
10.1.1	Item Response Theory (IRT) Estimation and Test Equating	90
10.1.2	Vertical Scaling	91
10.1.3	Composite Subject Knowledge Indices	92
10.1.4	Testing for Robustness to Equating Method	93
10.1.5	Draw a person test	94
10.1.6	Local content	94

10.2 Pupil Level Outcomes	95
11 Interpretation and Extensions of Results	106
11.1 The Bridge model	106
11.2 Heterogeneous effects	109
11.3 Income effect of the scholarship	112
11.3.1 Expenditures	112
11.3.2 Missed class	113
11.3.3 Sibling spillovers	114
11.4 Mechanisms and inputs in education production	115
11.4.1 Direct or proxy measurement of inputs	115
11.4.2 Heterogeneous effects	116
11.5 Variation in effects across academies	119
11.6 Effect on dispersion of test scores	120
11.7 Heterogeneity across teacher skill levels	120
11.8 Heterogeneity across class sizes	123
11.9 Vulnerability to manipulation and external validity	125
11.10 Cost Effectiveness	126
A Randomization Details	132
A.1 Round 1	132
A.2 Round 2	136
A.3 Selection Criteria	137
A.3.1 Sibling Factor	137
A.3.2 Academic Achievement Factor	137
A.3.3 Vulnerability Factor	138
B Endline Sampling	138

C	Assessment Development	140
D	Formal LATE decomposition	143
E	Derivation of LATE decomposition expression	145
F	Leave-one-out counterfactual predictions	147
	F.1 Leave-one-out academy means	148
	F.2 Empirical Bayes Shrinkage: Binomial with Beta prior	148
G	Construction of endogenous attendance variables	149
	G.1 2017 attendance classification	151
	G.2 2016 attendance classification	152
	G.3 2018 and 2019 attendance classification	153
H	Appendix Figures	155
I	Appendix Tables	155

1 Introduction

This document describes stage one analysis of an evaluation of Bridge International Academies in Kenya and describes plans for a second stage of analysis. This document will describe the research team’s data access, the results of a first stage of analysis, and decisions regarding sample restrictions and data processing for subsequent analyses.

Bridge operates several hundred private pre-primary and primary schools throughout the country. The present study will investigate a variety of dimensions of the Bridge model in Kenya. These include two main research questions regarding the impact of Bridge on test score outcomes: (1) Do the learning outcomes of students who receive a scholarship to attend a Bridge school improve (or worsen) compared to students who do not receive a scholarship to attend a Bridge school? (2) What mechanisms underlie the estimated impacts?¹ Below we provide a roadmap for this document.

Section 2 describes Bridge schools in Kenya – highlighting key differences between Bridge and other schooling options, as well as the active public discussion surrounding Bridge operations in the country.

Section 3 provides details of the scholarship lottery that we examine. The lottery took place at the end of 2015 and covered school fees for scholarship winners to attend any Bridge school of their choice in 2016 (later extended to include 2017 as well).

Section 4 outlines the data that was collected both to evaluate the effects of the scholarship lottery (and shed light on potential mechanisms that underlie any effects that we may find), and to better characterize the learning environment at Bridge schools.

Section 5 presents results describing characteristics of Bridge schools and the teachers

¹We note that this study differs in many respects to a recent evaluation in Liberia that also included Bridge International Academies (Romero et al, 2017). In Kenya Bridge operates its own schools, whereas in Liberia Bridge was part of a program where private operators contracted out their services into public schools. These operators, including Bridge, were required to work with existing teachers, administrators, and school facilities. In addition, the educational environments are quite different. Liberia had, and has, one of the weakest, lowest-performing education systems in Sub-Saharan Africa (coming off a decade of civil war), while Kenya’s educational system is one of the best in the region.

employed at those schools. This section will also compare these results to similar results for nearby schools.

Section 6 presents results on endline follow-up rates, differential attrition, and statistical balance of treatment on baseline covariates.

Section 7 describes interim results that have been obtained as part of a first stage of analysis. These results provide information that will be used to plan subsequent analyses. This information includes the first stage effect of treatment on school attendance, the effects of the scholarship on attendance at Bridge *at cost*, the effects of the scholarship on sibling school attendance, the effects of a brief scholarship extension that Bridge provided during the home-based endline survey on attendance and school switching, and the effects of the scholarship on grade progression.

Section 8 describes the goals of this document, including how it differs from pre-analysis plans. This section will explain why we chose to conduct the analysis in stages.

Section 9 presents the plan of analysis to estimate the causal effect of attending Bridge – using the scholarship lottery – on learning and other related outcomes. It also explains plans for sub-group analysis, e.g. by schooling level and by alternate outside educational options.

Section 10 describes the full set of outcomes that will be investigated. These outcomes include those that can be studied using the scholarship lottery as well as those that can be explored using observational data collected as part of the study or from secondary sources.

Finally, Section 11 explains the issues around the interpretation of the results and describes directions for extensions of the analysis.

2 Description of Bridge International Academies

Bridge operates several hundred private pre-primary and primary schools throughout Kenya.² The model used at Bridge schools to deliver education differs considerably from those em-

²Bridge also operates schools in India and Uganda, and is collaborating (to various degrees) with governments in Liberia and Nigeria.

ployed in other schools in this context. Reviewing these differences in educational/instructional approaches helps guide our analysis, especially those concerning mechanisms.

Bridge teachers use highly detailed lesson plans (sometimes referred to as scripts) in their instruction. These are created centrally and used by teachers via tablet e-readers. Bridge states that these lessons are designed to follow the national curriculum, but some stakeholders have raised concerns about how closely these lessons adhere to the curriculum. An open question is how teacher capabilities (and knowledge) interacts with this method. For instance, it is possible that the technology provides a scaffold for teachers and improves the pedagogy of less capable teachers. It could also be that it discourages individual teachers from targeting or personalizing instruction to better suit their classroom. Although teacher recruiting practices have changed, during our study Bridge recruited teachers who did not necessarily have the three years of teacher training college that is the norm in Kenyan public schools. This could lead to differences in teacher capabilities.

The tablet technology can also be used for monitoring both pupil and teacher performance. For instance, teachers can easily access historical pupil test score data on their tablets, potentially allowing them to better monitor student learning and progress. The tablet can also be used to monitor teachers (e.g. teacher attendance and teacher progress on lessons). Public schools and private schools in this context typically keep records on paper.

In addition to the use of technology, the Bridge model could differ in other ways. Bridge has developed its own textbooks and learning materials (such as student homework books) and aims to provide each student with their own set of books.³ This is in contrast with public schools where textbooks and other materials are often shared (World Bank 2013 and Uwezo 2012). If the quality and per-pupil availability of the Bridge textbooks differs from the alternatives in the market, then this could affect learning. There could also be differences in several other dimensions such as the actual instructional time (private school tend to have longer school days than government schools), class size, amount of homework assigned, and

³Some stakeholders argue that these materials are not well aligned to the Kenyan curriculum.

remedial instructional practices. During the period of this study (2016-2017), school fees were approximately 100 USD per pupil per year.⁴

3 The Scholarship Program

This section provides details on the scholarship program, includes information on recruitment and application procedures, and describes the randomization process that was used to select scholarship lottery winners.

3.1 Overview

The evaluation is centered on a scholarship program jointly sponsored by UnitedWeReach (UWR), an independent donor and Bridge itself. UWR provided funding for the scholarship program “in order to provide improved educational and life opportunities for the residents of those communities.”⁵ Children from all grades in pre-primary and primary school were eligible to apply for a scholarship to attend any of the 405 Bridge schools then operating throughout Kenya (see Figure 1). Ultimately the program allocated scholarships to approximately 10,000 students chosen by a lottery from among more than 25,000 applicants.

⁴It should be noted that Bridge costs per pupil exceed the tuition fees charged to families.

⁵UWR website: <http://unitedwereach.org/projects-scholarships/>

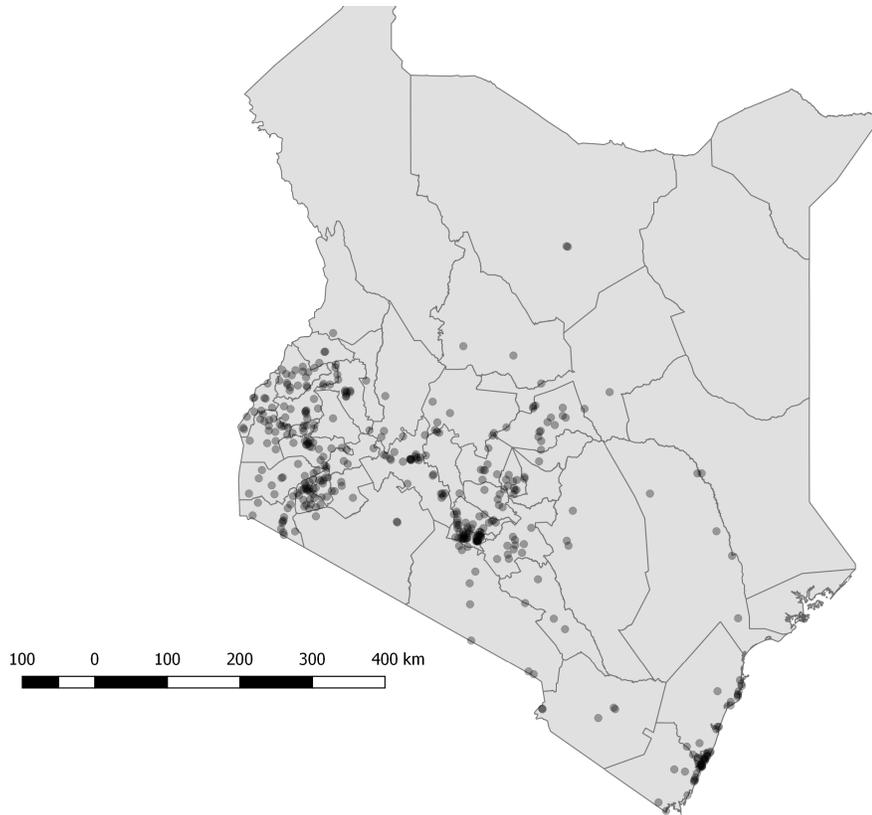


Figure 1: Location of 405 Bridge academies operating in 2015.

Initially, scholarships were provided for one academic year, starting in January 2016. However, in August 2016, scholarship recipients were informed scholarships would be renewed for the 2017 school year.⁶ These scholarships were available for pupils in any of the three pre-primary levels of Kenya’s schooling system⁷ as well as for pupils starting all but the last year of primary school.

3.2 Scholarship application recruitment and procedures

Bridge announced the scholarship program in November 2015 and disseminated information about the program through multiple channels. The program was advertised through posters

⁶As will be discussed later in Section 7.4, in January 2018 scholarships were extended for an additional one term for a subset of 2,266 pupils (the Kenyan school year corresponds roughly to the calendar year and consists of three terms of approximately three months).

⁷Preschool in Kenya consists of baby, nursery, and preunit classes. Children in each class tend to be three, four, and five years old, respectively.

and fliers distributed in communities around existing Bridge academies. Bridge called parents of former pupils and potential pupils Bridge had identified as part of its usual, rolling recruitment process. Bridge also sent out SMS messages to parents of currently enrolled Bridge pupils to encourage applications from among non-enrolled siblings, relatives, and neighbors. Finally, Bridge asked its academy managers⁸ to seek out additional potential applicants through community outreach.

Separately, staff from Innovations for Poverty Action (IPA) conducted door-to-door outreach in 100 randomly chosen Bridge academy locations to provide information on the scholarship program. IPA did not advocate for or otherwise encourage people to attend Bridge; however, if parents and pupils decided they would like to apply to the scholarship program, IPA provided them with assistance. Overall, IPA assisted with 2,612 applications while 27,357 applications were made through other channels for a total of 29,969 applications.

Applications for the scholarship program were accepted in November-December 2015. Parents of interested pupils were asked to complete an application form either online, through a call center operated by Bridge, or in person at a Bridge academy. The application form requested information on household demographics, basic socioeconomic status, the gender, age, and other information about the pupil, and the top two Bridge locations which they hoped to send their child. Additional details on the data collected as part of the application are described in the Section 4 below.

All pupils applying to the scholarship program who were not currently enrolled at a Bridge school were required to visit a Bridge academy to complete their application. The purpose of this in-person visit was twofold: (1) to reduce the likelihood of multiple applications for the same pupil or “phantom” applicants⁹, and (2) for pupils entering at the primary school

⁸Each Bridge academy has a manager who has responsibilities similar to that of a head teacher in other schools. These responsibilities include recruiting and community outreach, receiving payment, coordination with the central office, supervising teachers, and chairing meetings of the parent association.

⁹This phrase refers to the possibility that the open application process may have attracted some applications that were intended purely to exploit a perceived arbitrage opportunity. Members of the community (including academy managers) may not have understood that the scholarship would be paid directly to Bridge and non-transferable. If someone believed that the scholarship could be transferred to another student, then there would be a perceived incentive to submit a high volume of applications and resell the

level, to administer a Bridge placement exam to better determine what grade applicants would enter if they ended up matriculating to a Bridge school in the 2016 academic year.

3.3 Randomization

The broad contours of the randomization procedure are presented below (the finer details of this procedure are described in the Appendix A).

3.3.1 Defining sub-groups

The UWR scholarship program was open to all students – including those who, at the end of 2015, were enrolled at Bridge schools, government schools, other private schools, or those who were unenrolled. Probable outside options in the event of not winning a scholarship differed among students and treatment effects (if any) could possibly differ as well. Prior to randomization, scholarship applicants were classified by their probable outside schooling choice for the upcoming academic year.

The categorization of pupils was based on responses collected in the application form regarding current (2015) enrollment and planned (2016) enrollment in the event a scholarship was not awarded. From these responses two sub-groups were defined at the pre-primary school level and three sub-groups at the primary school level.

At the pre-primary school level, pupils were classified based on whether they were already attending a Bridge pre-primary school prior to 2016.

- Sub-group PP_{govt} consists of pupils who were not currently enrolled in a Bridge pre-primary school at the time of application (2015).
- Sub-group PP_{brig} consists of pupils who were currently enrolled in a Bridge pre-primary school at the time of application.

winning applications.

At the primary school level, pupils were categorized based on whether they were already attending a Bridge primary school or, if not, whether they planned to attend a government or another non-Bridge private school (including NGO or church administered) in 2016.

- Sub-group P_{govt} consists of pupils who were not attending a Bridge primary school at the time of application and who either attended a government primary school in 2015 or stated that they were planning to attend a government primary school in 2016.
- Sub-group P_{brig} consists of pupils who were attending a Bridge primary school at the time of application. This group also includes pupils who were matriculating from a Bridge pre-primary school and who stated they planned to remain at Bridge in the absence of winning a scholarship.
- Sub-group P_{priv} consists of pupils who were not attending a Bridge or a government primary school at the time of application and who were not planning to attend a Bridge or government primary school in 2016. This sub-group therefore consists of pupils who were likely to attend some low cost private school aside from Bridge.

3.3.2 Stratification

Each of the groups described above was stratified at the Bridge location and “grade group” level. In the application, families indicated the top two Bridge locations which they hoped to attend. Although students who received a scholarship were free to use their scholarship at any Bridge location with space, the applicants’ first choice was used as a stratifying variable for randomization. Grade group levels were defined as “Early Childhood” (Baby class, Nursery, and Preunit), “Lower Primary” (Standards 1-4), and “Upper Primary” (Standards 5-7).¹⁰

The number of scholarships assigned in each stratum depended on current enrollment at the Bridge Academy and the number of applicants to the Academy. Bridge wanted to

¹⁰The scholarship program also received applications for Standard 8. However, Bridge administered a separately scholarship program for Standard 8 students that as not randomized, so they are excluded from this study.

ensure that the share of scholarship students did not exceed 15 percent of total Academy enrollment. In addition, no more than 60 percent of applicants within a stratum were awarded scholarships. These constraints were modified slightly over time during multiple waves of assignment to reallocate scholarships as some applicants declined their offers (see the Appendix A for fuller details on how these constraints were adjusted).

3.3.3 Final randomization sample sizes

The randomization was done using random numbers generated by Stata. The sample sizes for each sub-sample are reported in the first row of Table 1.

Overall, this sample includes applicants to all 405 then-operating Bridge academies and all grades (aside from standard eight). 9 academies are not represented in the PP_{govt} and P_{govt} sample due to insufficient demand.¹¹ 10 academies are not represented in the PP_{brig} and P_{brig} sample due to insufficient demand. 201 academies are not represented in the P_{priv} sample due to insufficient demand.

All PP_{govt} and P_{govt} students were sampled for endline. Sub-samples were taken for the PP_{brig} , P_{brig} , and P_{priv} students. The sampling scheme is described in more detail in Appendix B. The sample sizes for each sub-sample are reported in the second row of Table 1.

3.4 Endline Sample

This sub-section describes further restrictions of the sample of pupils included in the endline sample that will be used for analysis.

As will be described later in Section 4, Lamu (N=47) and Garissa (N = 18) could not be visited at the time of the endline survey due to security concerns.

Initially, the research team had prepared for a small scale scholarship program involving approximately 2,000 pupils residing near a sub-sample of approximately 100 Bridge

¹¹These academies did have some applicants, but all applicants were in singleton randomization strata. See Appendix A for a specific information stratification.

academies. However, the announcement of the UWR scholarship program provided the research team with an opportunity to examine the effects of Bridge at a larger scale than would have been possible with the original, smaller research design. The evaluation of a large-scale, independent scholarship program afforded a larger sample size, but the decentralized nature of Bridge’s recruitment process in which academy managers recruited applicants meant the research team has less information on these applicants. In some cases, this process has generated data quality concerns. The research team is especially concerned that some applications may have been created in the hope of transferring the scholarship to other children. Although identities were supposed to be verified to redeem the scholarship, collusion with academy staff could have compromised that process. Illicit transfers of scholarship could compromise the fidelity of the randomized assignment and may also induce selective attrition. In some cases, when estimating the effects of the scholarship program on pupils, we will exclude certain applicants recruited through the large scale program where there are concerns the reliability of application data in order to ensure internal validity of the estimated effects. The large sample size means that it is possible to restrict the sample to the most reliable subset while not substantially reducing statistical power.

At endline, 420 applications were found to be duplicates. These pupils are dropped from the analysis.

Among samples of students who had not previously enrolled in Bridge (P_{govt} , PP_{govt} , and P_{priv}) and whose applications were received by UWR, two counties were flagged. One concern with the large scale scholarship program is that some applications may have been submitted in an effort to exploit a perceived arbitrage opportunity. A member of the community, or even an academy manager, could have seen an opportunity to submit several applications and then sell the right to use the winning applications. The application process included steps to verify identities, but it is possible that parents would have their child assume a new identity in order to use a scholarship, either without the academy manager’s knowledge or in collusion with the academy manager. Although the research team feels that the scope for fraud was

low, in order maintain the integrity of the overall experiment, some applications recruited by Bridge will be excluded from the tests described later in this document. Specifically, the county of Meru contained several academies with high volumes of applications and Nakuru had a low follow-up rate and a highly significant differential between treatment and control. We will exclude applications from UWR for these two counties for pupils who were not previously enrolled at Bridge schools.¹²

Anomalies in the county of Meru were identified prior to endline activity on the basis of high application volumes and low phone call follow-up rates. While the scholarship application was open, Bridge flagged one academy manager in Meru for receiving payments from families in exchange for scholarship applications. This same academy was found to have the highest volume of applications in the county. Several other academies in Meru recorded very high application volumes, as well. Meru represented 6.6 percent of applications, but the county represents only 3 percent of Kenya's population.

Nakuru was identified as anomalous by testing for heterogeneity across counties in the degree and statistical significance of differential attrition.¹³ The research team tested, for each county individually, whether the coefficient on the interaction of a county dummy and treatment was statistically significant (including strata controls). This test indicated that, among applications received through UWR, Nakuru county has a 13 percentage point higher attrition differential and the p-value of this difference is < 0.0001 . This result strongly rejects the hypothesis that the differential for the 247 applications received through UWR

¹²Although it is unlikely that this decision will substantially influence the overall estimate because these locations represent a small fraction of the overall sample, we will conduct sensitivity checks that include this sample to assess whether this decision significantly influences the results.

¹³This restriction of the sample is being made prior to observing endline outcome data. The resulting estimate on the restricted sample will be identified and inferential tests will be valid for the remaining subsample. To the extent that estimated effects are heterogeneous across counties, this choice may affect the external validity of the estimate with respect to the overall sample of students who applied to the scholarship. However, because the restriction excludes a relatively small portion of the total sample, it is unlikely that the populations are substantially different. By including applicants that were received through IPA in Nakuru, this county is still included in the estimated effect. This sample restriction only diminishes the weight placed on the county. Additionally, the analysis will include tests of heterogeneity in effects that will shed light on the possibility that the effect of Bridge varies substantially across regions. As such, the research team feels that any costs due to a reduction in external validity is compensated for by the benefit of increased transparency and confidence in results.

for Nakuru county is the same as the remaining sample. Some other counties have larger differentials in magnitude, but these tend to be smaller counties.¹⁴

In these two counties, the study will restrict the analysis to those students who entered the scholarship application process through IPA staff. The IPA sample has a high, balanced follow-up rate and, when restricting to these data, the anomalies in these counties disappear. Note that this restriction applies only to the samples of students who were new to Bridge in 2015. Samples of students who were enrolled in Bridge at the time of the scholarship offer have high follow-up rates and attrition is balanced between treatment and control.

The resulting sample sizes (ignoring, for now, follow-up rates) are given in the bottom row of Table 1.

Table 1: Construction of final sample

	PP_{gov}	P_{gov}	PP_{brig}	P_{brig}	P_{priv}
Randomized sample	5,145	6,505	2,925	9,356	2,293
Selected for endline	5,057	6,250	435	1,416	1,436
Excluding Lamu and Garissa	5,044	6,231	435	1,393	1,434
Removing duplicates observed at endline	4,835	6,059	433	1,390	1,395
Removing UWR applicants in Meru and Nakuru	4,411	5,267	433	1,390	1,305
Final Sample	4,411	5,267	433	1,390	1,305

3.5 Siblings sample

Siblings can be partially identified in the application files using a combination of contact phone numbers and information on the relationship of the contact to the pupil. All identified siblings of students sampled for endline using the procedures described above were also included in the endline. 661 P_{brig} pupils and 157 PP_{brig} pupils were selected because they were siblings of pupils that were sampled for endline through the random sampling scheme described above.

¹⁴For example, this same test shows that Baringo has a 31 percentage point large attrition differential, but this is estimated from a sample of 23 applications. The test statistic is only 2.81. Given the very small number of applications from Baringo, the research team has chosen to include these applications in the study.

4 Data

This section provides details on the data that were collected from pupils in the study and their caretakers, as well as school level data collected separately.

A timeline of survey activities is provided in Table. 2

Table 2: Timeline

Activity	Month	Year
Applications received for scholarship	November	2015
Scholarship assignment and notification of award	December	2015
Phone call survey (2016)	June	2016
Local education market survey	October	2016
Phone call survey (2017)	July	2017
Endline survey	November	2017
Phone call survey (2019)	June	2019

4.1 Pupil Level Data

Data on scholarship program applicants were collected at several stages over the course of the experiment. Baseline data were collected at the application stage. Those students recruited by IPA also took part in more extensive baseline data collection, including cognitive and non-cognitive tests. In 2016 and 2017, phone call tracking surveys were conducted with the PP_{govt} and P_{govt} samples. Finally, from November 2018 - March 2019, endline data was collected for all PP_{govt} , P_{govt} and P_{priv} students and a randomly selected sub-sample of PP_{brig} and P_{brig} students.

4.1.1 Baseline Data

In addition to location preference data, the application collected data on all applicants on the following characteristics:

- Current education status (unenrolled, Bridge, government school, non-Bridge private school)

- Future (2016) education plans (unenrolled, Bridge, government school, non-Bridge private school)
- Current grade in 2015
- Planned grade in 2016
- Gender
- Disability status (vision, mobility, hearing)
- Household income (Kenyan shillings, monthly)
- Number of children and adults in the household
- Household asset (TV, radio, cell phone, etc.)
- Whether household owns or rents their home
- Material of household's roof (mud, timber, bricks, Canes or planks, cement, mabati [metal sheet])
- Material of household's walls (mud, timber, bricks, Canes or planks, cement, mabati [metal sheet])
- Presence of a latrine in the household
- Household access to electricity
- Term 3 end of term tests in English and math (self-reported)
- Information on whether parents are alive
- Information on the literacy and education of the caregiver
- Caregiver employment status and occupation

More extensive baseline data were collected for a subset of 1,931 applicants from the sample interviewed by IPA Staff . These data include a test of self-regulation, an adapted version of the Malawi Developmental Assessment Tool (MDAT), Raven's matrices, Uwezo's early primary test, and Bridge's placement tests.

4.1.2 Midline Phone Call Data

Both phone call surveys were conducted with only those students in the PP_{govt} and P_{govt} samples. At the time these surveys were conducted, the research team planned to follow only this group at endline. The primary purpose of the phone call survey was to maintain regular contact with the study participants to ensure high quality contact data would be available for conducting endline data collection.

2016 Phone call survey

The 2016 phone call survey was conducted in October. The survey collected data on the following:

- School enrollment (government, Bridge, non-Bridge private, non-enrollment)
- Grade reached
- Contact information of parents, guardians, and other family and friends.
- A list of nearby school options

The research team used the data on school choice to examine the first stage effect of the scholarship on Bridge attendance, estimate the share of students in the control group who were attending Bridge, and study the choice of educational environment of both scholarship recipients and non-recipients. Understanding the first stage effect on Bridge attendance was essential to conducting power analysis and to determine the necessary sample size needed at endline.¹⁵ These data were also used to develop a better understanding of the counterfactual against which Bridge attendance would be compared.

¹⁵Because the scholarship program included students dispersed across 44 out of 47 Kenyan counties, the research team was concerned about the feasibility of following up with the full sample. Given the first stage effect of the scholarship on Bridge attendance between 0.3-0.4 observed in this phone call survey, the research team determined that following up with only a subsample would yield an unacceptable reduction in the precision of the estimates.

2017 Phone Call Survey

The purpose of the 2017 phone call survey was to maintain contact with participants in the study. A concern, prior to the start of endline activities, was that students could have moved out of areas where they were initially observed at baseline. Because the endline survey was in-person, such moves would have increased the logistical complexity of the survey, so this information was important to accurately forecast costs. The survey was also intended to provide information on the schools attended by study participants in 2017 and to observe whether the first stage effect on Bridge attendance was still sufficiently large to ensure adequate precision for endline estimates.

The data collected in this survey primarily consisted of checking information regarding the identity of the applicant and confirming and revising contact data of caregivers.

Follow-up rates and estimates of differential attrition are reported in Table G1.

4.1.3 Endline Data

The endline survey consisted of a parent survey, a pupil survey, and pupil assessments in specific academic subjects. The parent survey collected data on family characteristics, educational expenditures and school attendance of all household members younger than 18, parental involvement in the pupil's education, and satisfaction with educational services. The parent survey also collected rudimentary income data for all households. The pupil survey collected data on the pupil's school and classroom experience, opinions about school and teaching, pupil's absenteeism and effort, and a brief module on opinions on matters related to civics. The survey also included tests of cognitive and non-cognitive skills.

Endline survey activities were done in person, generally at the student's home. In some cases, where there were security concerns, parents and students were invited to come to a central testing center. These circumstances are described below.

Security Constraints

Lamu and Garissa were excluded from endline data collection activities due to security concerns. Other counties were visited late during endline activities. The following counties were on IPA's security watch list at the time endline activities began: Baringo, Elgeyo-Marakwet, Mandera, Marakwet, Marsabit, Moyale, Samburu, Tana River, Turkana, Wajir, and West Pokot.

Additionally, at the start of endline activities, the survey team was unable to visit Trans Nzoia due to an outbreak of Marburg Virus. The research team conducted some interviews and testing in some security risk counties by setting up testing centers and inviting students to these centers. This approach yielded lower follow-up rates than home visits. As the security situation improved, home visits were allowed in most of these areas and the overall follow-up rate in these locations improved. Given the potential for endogeneity in testing environments, results will be reported that drop those Bridge locations where centers were employed, as a robustness check.

Several counties, including Nairobi, at times also could not be surveyed due to security issues related to national elections.

Data overview

One empirical question is the effect of Bridge attendance on pupil performance on tests of academic subject knowledge. Students received grade appropriate subject level tests that are based on the Kenyan national curriculum. Students are tested on material appropriate to the grade which they completed at the end of the most recent school year, 2017. Consultants with teaching and curriculum development skills and knowledge were engaged to design and develop these subject level tests over a 3 month period with the tests being pre-tested with non-sample pupils. Additional information, including estimates of the reliability of each test instrument are reported in appendix C.

Students in any level of ECD were tested in two subjects:

1. Reading (in English and Kiswahili)
2. Math and Patterns

Students in Standards 1 – 8 were tested in 4 subjects

1. Mathematics
2. English
3. Kiswahili
4. Social Studies and Science

Endline also consisted of parent and pupil surveys. These data collected information on pupil school of attendance, current grade, as well as several other outcomes that will be explored in the analysis. For more details on the data available in the endline survey, see Section 10.2.

4.1.4 2019 Phone call follow-up

Between June and November 2019, caregivers of pupils in the home-based endline survey were contacted by phone to conduct a brief tracking survey. Data collected in the survey included contact information, information on the pupil's grade in 2019, school attendance status (including the type of school attended), the amount of school missed due to fees, and information about KCPE scores. For pupils who had not taken the KCPE exam, caregivers were asked to report expectations about results of the KCPE exam, including whether the caregiver expected that the pupil would take the exam and the score that the caregiver expected the pupil to receive. For pupils who had already taken the KCPE exam, caregivers were asked to report the pupil's KCPE score and the location at which the pupil took the KCPE exam.

KCPE results for students taking the KCPE exam in 2019 were released in November. Students enrolled in Standard 8 according to the tracking survey in 2019 were recontacted after KCPE scores were released to collect results on their KCPE performance.

4.2 School Level Data

In addition to data on scholarship applicants, a survey of schools was conducted in 2016 to collect information on the local education options available to families in the areas around Bridge academies. This activity is described below

4.2.1 Local Education Environment Survey

The local education environment survey was conducted to shed light on the educational choices available to students participating in the scholarship program. The survey was conducted from September 13, 2016 until November 10, 2016. The activity was originally intended to collect information on the local education options for all of the 405 Bridge then-active Bridge academies. However, the activity was cancelled due to cost considerations.

16

The local education environment survey identified schools for the survey using a snowball sampling scheme. The typical respondent for the survey was a head teacher or academy manager. Field officers surveyed the Bridge academy first. At the conclusion of the survey, the respondent was asked to identify other schools in the area. These schools were then surveyed and the respondent at those schools was asked to identify other schools in the area. The procedure ended when sampled respondents did not identify new schools within 3 km of the original Bridge academy.

In addition to those schools identified by the snowball sampling scheme, schools attended by applicants to the scholarship program who applied to the Bridge location were also sampled.¹⁷

¹⁶Costs exceeded expectations because the number of private early childhood education options was much larger than expected.

¹⁷These schools were identified in the 2016 phone call survey. The school names were not standardized prior to the survey and did not have unique identification codes. At the time of the survey, field officers had a list of all reported schools from the phone call survey, including multiple spellings of the same school. It is therefore not possible to precisely identify those schools that were identified by using the snowball sampling scheme and which were specifically mentioned by Bridge students. However, using a fuzzy matching algorithm, it is possible to identify which survey schools most closely resemble the names of schools reported in the phone

Before the activity was concluded, a total of 3,476 schools were identified. 3,426 respondents consented to the survey. 113 of these schools were Bridge academies. A total of 125 markets were surveyed. 78 of these were determined to be completely surveyed in that the sampling scheme had reached a terminating point, having identified all schools within the radius. Markets that were not indicated as complete have on average approximately 30 schools, compared to markets that were indicated as complete which have on average 40 schools.

The coordinates of the schools are mapped in Figure 2.

This survey also collected data on 34,653 teachers, including 1,337 teachers at Bridge academies. The data are limited to the grades taught, years of experience (at the school), and highest level of education completed. A question on the total number of years of experience teaching and TSC certification was added later in the survey, but coverage for these variables is low (882 non-missing observations).

While the selection of Bridge academies into this sample was not randomized, the sampled academies represents 34.46 percent of the overall sample of applicants surveyed at endline.

5 Comparison between Bridge and Nearby Schools

This section describes characteristics of Bridge schools and nearby schools as observed in data collected as part of this project.

These data reveal several contrasts between Bridge schools and nearby government schools. Compared to nearby government schools, Bridge schools have lower pupil-teacher ratios, operate longer school days, have higher fees, and employ teachers with less education.

The local education environment (LEE) survey provides information on school fees, pupil-teacher ratios, length of school day, quality assurance, formal registration, and the average KCPE score of students. These data also include information on whether the school participates in the Tusome or Tayari programs, government initiatives that were ongoing during the call survey.

evaluation and aimed at improving pedagogy in lower primary and pre-primary respectively. These programs may be effective in improving the quality of public schools and may therefore their presence in schools could provide important context in interpreting heterogeneity in the effects of the scholarship program across grade levels that received or did not receive these programs. The survey also includes information on teachers at schools including their highest level of education and experience teaching.¹⁸

The results reported here are for the 125 academies (and the schools in their surrounding area) for which the activity was started.¹⁹ See Section 4.2.1 for details on this survey activity.

Pupil-teacher ratios are computed by dividing the total number of students enrolled in pre-primary grades and dividing by the total number of teachers reported to ever teach pre-primary grades.

The survey asked head teachers whether the school had been visited by any quality assurance officers in the past year. The Ministry of Education may enter schools for quality assurance purposes, but, in the case of Bridge and other private schools, such visits could also reflect internal quality assurance inspections.

The survey asked head teachers whether the school was taking part in the Tayari program. The Tayari program is a joint initiative of the government of Kenya and Research Triangle Institute (RTI) International that provides government and low cost early childhood education centers with support (in the form of training and the development of teaching and learning materials) to develop literacy, numeracy, and socio-emotional skills to prepare children for primary school.

The survey also asked head teachers about their participation in the Tusome program, a national literacy program for Standard 1 through 3. The program is also a joint initiative of the government of Kenyan and RTI. The program provides materials and training to teachers building on lessons learned from the Primary Mathematics and Reading (PRIMR) Initiative, which was found to have positive effects on literacy (Piper et al. (2014)).

¹⁸Additional data were collected as part of this survey but have not yet been analyzed.

¹⁹As described in Section 4.2.1, the snowball sampling scheme was only completed in 78 locations.

Characteristics of individual teachers such as level of education and the amount of time that the teacher has been employed at the school are reported by the head teacher/academy manager. The data are not collected through an interview with the individual teachers. The survey collected the number of years the teacher had been teaching at the school. Data on total teaching experience were not collected. The survey also collected data on highest education level attained by the teacher. Responses included secondary school, certificate, diploma, bachelors, and masters. Certificates and diplomas are common professional qualifications and diplomas tend to require more time than certificates. Although it is likely that many of the reported certificates and diplomas refer to teacher training certificates, it is possible that these include other types of professional certificates.

The results of the survey are reported separately for pre-primary, lower primary, and upper primary programs. Table 3 shows the results for pre-primary schools. Pupil-teacher ratios in Bridge schools are lower than both government and private pre-primary schools in the SDI survey²⁰, and the length of the school day is longer. At government schools, the length of a pre-primary school day is 4.4 hours on average, compared to 7.5 hours at Bridge schools. At private schools, the length of the school day 5.9 hours. The distribution of school days for private pre-primary schools appears to be bi-modal²¹, indicating that many private pre-primary programs are half-day programs. Table 3 also indicates that government pre-primary programs charge fees, although these fees are typically much less than Bridge.

Table 4 and Table 5 show results for lower and upper primary students. Across all grade groups, Bridge classrooms have lower pupil-teacher ratios than government schools, although in upper primary the government pupil-teacher ratio is lower than in lower primary, so the difference in pupil-teacher ratios between Bridge and government schools is not as stark at the lower primary level. Bridge schools also have longer school days than government schools, although this gap narrows in upper primary. For younger students, Bridge also has

²⁰Note that because the SDI survey took place in 2012, it is possible that some Bridge schools were included in this survey. However, it is unlikely that Bridge schools account for a large portion of the private schools in the SDI survey.

²¹See Appendix Figure 3.

lower pupil-teacher ratios than other private schools. Bridge has longer school days than other private schools for younger students. However, for upper primary students, private schools have similar pupil-teacher ratios and the length of the school day is approximately the same.

Table 3: Comparison Bridge and nearby schools - Pre-primary level

	Bridge (1)	Government (2)	Private (3)
<i>School Characteristics</i>			
Pupil-teacher ratio	14	44	24
Fees (kes)	6,766	1,699	7,527
Length of school day (hrs)	7.5	4.4	5.9
Supervisor conducts classroom observations	0.72	0.44	0.45
School visited by QA officer last year	0.66	0.85	0.45
School registered with the TSC	0.02	0.94	0.01
School has a KNEC code	0.15	0.94	0.23
School is in Tayari program	0.04	0.06	0.05
N schools	116	828	2290
<i>Teacher Characteristics</i>			
Female	0.85	0.94	0.97
<i>Highest level education</i>			
Secondary school	0.38	0.12	0.19
Certificate	0.48	0.53	0.59
Diploma	0.14	0.32	0.21
Bachelor or Master	0.00	0.01	0.01
<i>Experience/Employee tenure</i>			
Years teaching at this school	3.3	5.1	2.8
First year teaching at this school	0.26	0.16	0.33
N Teachers	308	2,067	6,433

Notes: This table shows results from the local education environment (LEE) survey of schools in the areas around a sub-sample of Bridge academies. Characteristics of individual teachers are reported by the head teacher/academy manager. Pupil-teacher ratios are computed by dividing the total number of students enrolled in pre-primary grades and dividing by the total number of teachers reported to ever teach pre-primary grades. “School visited by QA officer last year” indicates that any quality assurance officer visited the school. QA officers might be government inspectors, but, in the case of Bridge, could include internal QA officers. The Tayari program is a joint initiative of the government of Kenya and RTI that provides government and low cost early childhood education with support (in the form of training and the development of teaching and learning materials) to develop literacy, numeracy, and socio-emotional skills.

Table 4: Comparison Bridge and nearby schools - Lower primary level

	Bridge (1)	Government (2)	Private (3)
School Characteristics			
Pupil-teacher ratio	22	58	23
Fees (kes)	7,687	0	9,694
Length of school day (hrs)	8.4	5.8	7.4
Supervisor conducts classroom observations	0.72	0.43	0.43
School visited by QA officer last year	0.66	0.87	0.49
School registered with the TSC	0.02	0.96	0.02
School has a KNEC code	0.15	0.96	0.31
School is part of Tusome program	0.33	0.98	0.48
N schools	116	862	1754
Teacher Characteristics			
Female	0.66	0.86	0.81
<i>Highest level education</i>			
Secondary school	0.39	0.08	0.19
Certificate	0.52	0.48	0.63
Diploma	0.08	0.29	0.16
Bachelor or Master	0.01	0.15	0.01
<i>Experience/Employee tenure</i>			
Years teaching at this school	2.8	7.5	2.6
First year teaching at this school	0.35	0.09	0.36
N Teachers	336	3,139	3,972

Notes: This table shows results from the local education environment (LEE) survey of schools in the areas around a sub-sample of Bridge academies. These results restrict to schools that have a Standard 1 classroom. Characteristics of individual teachers are reported by the head teacher/academy manager.

Table 5: Comparison Bridge and nearby schools - Upper primary level

	Bridge (1)	Government (2)	Private (3)
<i>School Characteristics</i>			
Pupil-teacher ratio	10	21	11
Fees (kes)	8,793	0	13,319
Length of school day (hrs)	8.9	8.0	8.7
Supervisor conducts classroom observations	0.71	0.43	0.41
School visited by QA officer last year	0.66	0.87	0.58
School registered with the TSC	0.02	0.97	0.03
School has a KNEC code	0.15	0.99	0.66
Average KCPE score	257	248	288
N schools	63	843	623
<i>Teacher Characteristics</i>			
Female	0.46	0.50	0.32
<i>Highest level education</i>			
Secondary school	0.25	0.06	0.15
Certificate	0.63	0.40	0.68
Diploma	0.10	0.28	0.12
Bachelor or Master	0.02	0.26	0.06
<i>Experience/Employee tenure</i>			
Years teaching at this school	2.8	6.1	3.2
First year teaching at this school	0.38	0.15	0.33
N Teachers	514	8,124	4,635

Notes: This table shows results from the local education environment (LEE) survey of schools in the areas around a sub-sample of Bridge academies. These results restrict to schools that have either a Standard 7 or Standard 8 classroom. Characteristics of individual teachers are reported by the head teacher/academy manager.

The data indicate that government schools do not charge fees, in accordance with Kenya’s policy of free primary education (FPE); however, caregiver responses in other surveys will show that caregivers report paying substantial fees. The fees at other private schools are, on average, significantly higher at other private school than at Bridge.

At all grade levels, Bridge teachers appear to have fewer education credentials. Bridge teachers are more likely to have only completed secondary school or achieved a teaching certificate, while government teachers are more likely to have received a teaching diploma or hold an advanced degree. Bridge teachers at all grade levels appear to have been teaching for less time at their schools.

6 Endline Attrition and Covariate Balance

This section discusses the study’s follow-up rates at endline and differential attrition between treated and control students. The section begins by presenting overall attrition results for the full sample and subsamples. Sub-samples with higher rates of follow-up are also discussed. Then, results showing covariate balance on treatment conditional on follow-up are presented to assess the likelihood of selective attrition in any form. Finally, the estimation of attrition bounds is discussed.

Table 6 shows follow-up rates for the full sample and individual sub-samples. Follow-up rates for the PP_{govt} and P_{govt} samples are 85.9 and 87.0 respectively. Pooled, the PP_{govt} and P_{govt} samples have a follow-up rate of 86.5 percent and scholarship winners are approximately 1.2 percentage points more likely to be followed at endline, although this effect is only marginally significant. For pupils in the PP_{brig} and P_{brig} samples (pupils who were in Bridge at the time of the scholarship program), the follow-up rate is 87.3 percent, and treated units are approximately 1.6 percentage points more likely to be followed at endline. The P_{priv} sample (students who were likely to enroll in non-Bridge private schools in the absence of a scholarship), follow-up is lower at 75.8 percent, and there is some evidence that pupils

who received a scholarship are 4.8 percentage points more likely to be followed at endline. Overall, the follow-up rate for the study was 85.9 percent and scholarship winners were approximately 1.4 percentage points more likely to be followed at endline.

Sub-samples have been identified with higher follow-up rates to minimize the risk of selective attrition contaminating the estimated effects from the study. The research team will examine results on these sub-samples in order to assess the potential for biases arising from selective attrition.

The sub-samples were identified by observing that, at baseline, some students had interacted more than others with either IPA staff or Bridge prior to applying to the scholarship. Pupils are included in the “baseline contact” sample if one of the following conditions is satisfied:

- Pupils enrolled in Bridge at the time of the application
- Pupils who applied to the scholarship program with the assistance of IPA staff
- Pupils who appeared in a Bridge academy prior to assignment to be measured for a uniform
- Pupils who had a sibling enrolled in Bridge at the time of the scholarship program

As expected, pupils who had previously interacted either with IPA or Bridge staff before the scholarship program were more likely to participate in endline activities and the differential in follow-up between scholarship recipients and non-recipients is smaller.

The study will also report results that restrict to the sub-sample of students who applied through IPA. This sub-sample has a very high follow-up rate and no evidence of differential attrition. However, the sample size is too small for some of the analyses that are proposed below.

As a first step in the analysis of the effect of treatment on endline results, the research team will compare the treatment effect²² observed in the baseline contact sub-sample and the full sample. Out of an abundance of caution, if the absolute value of the test statistic on the

²²For a discussion of the reduced form treatment effect, see Section 9.1.

Table 6: Follow-up and differential attrition at endline

Sample	Full Sample			“High Contact” sample			IPA sample		
	Mean (1)	Diff. (2)	<i>N</i> (3)	Mean (4)	Diff. (5)	<i>N</i> (6)	Mean (7)	Diff. (8)	<i>N</i> (9)
PP_{gouv}	0.859	0.010 (0.010)	4,411	0.877	-0.008 (0.012)	3,098	0.889	-0.015 (0.023)	802
P_{gouv}	0.870	0.015 (0.009)	5,267	0.893	0.000 (0.011)	3,535	0.917	0.001 (0.018)	1,061
$PP_{gouv} + P_{gouv}$	0.865	0.012* (0.007)	9,678	0.885	-0.004 (0.008)	6,633	0.905	-0.006 (0.014)	1,863
PP_{brig}	0.896	0.039 (0.026)	433	0.896	0.039 (0.026)	433	1.000	- -	7
P_{brig}	0.866	0.009 (0.017)	1,390	0.866	0.009 (0.017)	1,390	0.850	0.160 (0.205)	20
$PP_{brig} + P_{brig}$	0.873	0.016 (0.015)	1,823	0.873	0.016 (0.015)	1,823	0.889	0.129 (0.164)	27
P_{priv}	0.758	0.048** (0.023)	1,305	0.772	0.063** (0.029)	869	0.793	0.015 (0.069)	198
Full Sample	0.859	0.014*** (0.006)	13,616	0.876	0.003 (0.007)	10,135	0.894	-0.003 (0.014)	2,095

Notes: Standard errors are reported in parentheses. Each coefficient estimate of differential attrition comes from a regression of an indicator for endline test score follow-up for the sample listed on the left hand column. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *gouv*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. The Full sample refers to the entire sample of students to be included in the analysis. This study will also explore results with a “High contact” sample of students who were observed either in Bridge data or interacted with IPA staff prior to scholarship assignment. The IPA recruited sample refers to just those applicants who were applied to the study through IPA. ***, **, and * indicate significance at 1%, 5%, and 10%.

coefficient for the interaction of treatment with a indicator for baseline contact individuals is greater than 1.15,²³ the research team would encourage readers to focus on the sub-sample with lower attrition: the baseline contact sample.

Tables 8 - 11 present results of covariate balance on treatment for each of the samples PP_{govt} , P_{govt} , PP_{brig} , P_{brig} , and P_{priv} samples. Each table shows results separately for the full sample and for the full sample conditional on following the pupil at endline. The covariates included in the balance test are from the baseline application. “Term 3 end-of-term” test scores in English and math are self-reported measures.²⁴ The first two columns show the results for the full sample. All tables include an F-test of joint significance. The F-test includes only those pupils who have non-missing values for all covariates. Because many pre-primary students do not have baseline subject knowledge in English and math, the table reports a test of joint significance of all variables excluding these tests.

While there is not evidence of selective attrition, it is still possible that differential attrition has resulted in unobserved differences between the treatment and control groups conditional on follow-up. To account for this possibility, the results from Lee bounds will be reported for the main effects of the experiment. However, the research team feels that given the limited evidence of selective attrition, Lee bounds are probably too extreme. The study will also estimate bounds within cells of baseline characteristics. In this case, the sample will be divided into quintiles of the baseline test score²⁵ distribution using and the distribution of the treatment group endline outcome will be truncated as in Lee bounds within each cell quintile separately.²⁶

²³1.15 corresponds approximately to $\alpha = 0.3$, or a 30 percent chance of a false rejection of the null hypothesis of equal effects.

²⁴These are not results from a standardized test. T3ET tests are designed locally by schools. Preliminary evidence suggests that these tests are predictive of endline performance.

²⁵The test score used to form these quintiles will be the reported Term 3 end-of-term math test score.

²⁶It is possible that, for some sub-samples the relationship between treatment and follow-up will not have the same sign within each quintile cell. In this case, more granular cells will be formed in order to ensure that, within all cells, the sign of the relationship between treatment and follow-up is the same.

Table 7: Covariate balance on treatment - PP_{govt} sample

	Full Sample			Observed at Endline		
	Mean (1)	Diff. (2)	N (3)	Mean (4)	Diff. (5)	N (6)
T3ET English Score (Pct)	0.764	0.006 (0.012)	1,732	0.759	0.010 (0.013)	1,478
T3ET Math Score (Pct)	0.771	0.011 (0.012)	1,733	0.767	0.015 (0.013)	1,479
Female	0.486	0.005 (0.015)	4,847	0.488	-0.001 (0.016)	4,120
Student has disability	0.007	-0.003 (0.002)	4,847	0.008	-0.004* (0.002)	4,120
Has sibling in Bridge	0.142	-0.009 (0.010)	4,847	0.154	-0.007 (0.011)	4,120
Mother is deceased	0.056	0.001 (0.006)	4,055	0.050	0.004 (0.007)	3,417
Father is deceased	0.192	0.017 (0.011)	4,055	0.189	0.016 (0.012)	3,417
<i>Caregiver Employment</i>						
Unemployed	0.224	-0.006 (0.011)	4,847	0.221	0.002 (0.013)	4,120
Worked in last 7 days	0.632	0.002 (0.013)	4,847	0.633	0.000 (0.015)	4,120
<i>Caregiver Occupation</i>						
Small business	0.175	0.000 (0.011)	4,847	0.178	-0.003 (0.012)	4,120
Agriculture	0.099	-0.010 (0.008)	4,847	0.110	-0.012 (0.009)	4,120
Casual labor	0.411	0.012 (0.014)	4,847	0.404	0.009 (0.015)	4,120
<i>Other household characteristics</i>						
HH Income (000s Ksh)	46.768	-0.251 (1.436)	4,847	47.292	-0.015 (1.612)	4,120
Caregiver can read	0.889	0.013 (0.008)	4,847	0.887	0.013 (0.009)	4,120
HH owns home	0.625	0.012 (0.012)	4,055	0.609	0.004 (0.014)	3,417
HH has electricity	0.605	-0.003 (0.012)	4,847	0.608	-0.004 (0.013)	4,120
HH has latrine	0.294	-0.013 (0.012)	4,847	0.292	-0.013 (0.013)	4,120
HH floor is dirt	0.544	0.001 (0.013)	4,847	0.554	0.003 (0.014)	4,120
Joint F stat		0.977			1.124	
p-value		0.484			0.323	
Joint F stat (excl. T3ET)		0.881			0.890	
p-value		0.592			0.581	

Notes: Standard errors are reported in parentheses. Each coefficient estimate in Columns 2 and 4 comes from a regression of the covariates listed on the left-hand column on treatment and strata dummies. Columns 1 and 2 show results for the full sample. An F-test of joint significance is provided at the bottom of the table. This test includes only those students who have non-missing values for all variables. The bottom two rows report a joint test of significance that excludes T3ET scores. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 8: Covariate balance on treatment - P_{govt} sample

	Full Sample			Observed at Endline		
	Mean (1)	Diff. (2)	N (3)	Mean (4)	Diff. (5)	N (6)
T3ET English Score (Pct)	0.654	-0.004 (0.005)	5,585	0.653	-0.008 (0.005)	4,754
T3ET Math Score (Pct)	0.652	0.003 (0.005)	5,585	0.650	0.005 (0.005)	4,754
Female	0.490	0.031*** (0.013)	6,078	0.493	0.027* (0.015)	5,203
Student has disability	0.007	-0.003 (0.002)	6,078	0.007	-0.002 (0.002)	5,203
Has sibling in Bridge	0.082	0.003 (0.007)	6,078	0.092	-0.004 (0.008)	5,203
Mother is deceased	0.090	-0.004 (0.008)	5,023	0.089	-0.003 (0.009)	4,236
Father is deceased	0.247	0.007 (0.011)	5,023	0.248	0.007 (0.012)	4,236
<i>Caregiver Employment</i>						
Unemployed	0.186	0.006 (0.010)	6,078	0.182	0.010 (0.010)	5,203
Worked in last 7 days	0.670	-0.007 (0.012)	6,078	0.668	-0.010 (0.013)	5,203
<i>Caregiver Occupation</i>						
Small business	0.185	-0.003 (0.010)	6,078	0.178	-0.002 (0.010)	5,203
Agriculture	0.151	-0.013 (0.008)	6,078	0.160	-0.013 (0.009)	5,203
Casual labor	0.405	0.008 (0.012)	6,078	0.407	0.005 (0.013)	5,203
<i>Other household characteristics</i>						
HH Income (000s Ksh)	42.150	1.040 (1.168)	6,078	42.220	0.978 (1.297)	5,203
Caregiver can read	0.871	0.003 (0.008)	6,078	0.866	0.007 (0.009)	5,203
HH owns home	0.519	0.008 (0.011)	5,023	0.499	0.008 (0.013)	4,236
HH has electricity	0.666	-0.013 (0.011)	6,078	0.656	-0.005 (0.012)	5,203
HH has latrine	0.279	0.011 (0.011)	6,078	0.273	0.014 (0.012)	5,203
HH floor is dirt	0.645	-0.014 (0.011)	6,078	0.653	-0.013 (0.013)	5,203
Joint F stat		1.031			0.915	
p-value		0.420			0.559	

Notes: Standard errors are reported in parentheses. Each coefficient estimate in Columns 2 and 4 comes from a regression of the covariates listed on the left-hand column on treatment and strata dummies. Columns 1 and 2 show results for the full sample. An F-test of joint significance is provided at the bottom of the table. This test includes only those students who have non-missing values for all variables. The bottom two rows report a joint test of significance that excludes T3ET scores. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 9: Covariate balance on treatment - PP_{brig} sample

	Full Sample			Observed at Endline		
	Mean (1)	Diff. (2)	N (3)	Mean (4)	Diff. (5)	N (6)
T3ET English Score (Pct)	0.834	-0.015 (0.018)	412	0.830	-0.013 (0.019)	368
T3ET Math Score (Pct)	0.868	-0.017 (0.018)	412	0.867	-0.021 (0.019)	368
Female	0.471	0.015 (0.050)	433	0.469	0.016 (0.053)	388
Student has disability	0.000	0.009 (0.007)	433	0.000	0.011 (0.008)	388
Has sibling in Bridge	0.420	-0.051 (0.047)	433	0.458	-0.087* (0.050)	388
Mother is deceased	0.040	0.006 (0.019)	430	0.044	0.005 (0.021)	385
Father is deceased	0.151	-0.013 (0.029)	430	0.159	-0.007 (0.032)	385
<i>Caregiver Employment</i>						
Unemployed	0.246	0.019 (0.039)	433	0.231	0.013 (0.041)	388
Worked in last 7 days	0.601	-0.011 (0.046)	433	0.606	-0.018 (0.050)	388
<i>Caregiver Occupation</i>						
Small business	0.116	0.046 (0.033)	433	0.113	0.050 (0.036)	388
Agriculture	0.076	0.015 (0.024)	433	0.083	0.015 (0.026)	388
Casual labor	0.411	-0.055 (0.047)	433	0.422	-0.065 (0.050)	388
<i>Other household characteristics</i>						
HH Income (000s Ksh)	48.930	-5.875 (4.648)	433	48.296	-4.649 (5.009)	388
Caregiver can read	0.905	-0.005 (0.028)	433	0.903	0.003 (0.029)	388
HH owns home	0.688	-0.026 (0.036)	430	0.666	0.002 (0.039)	385
HH has electricity	0.573	0.055 (0.044)	433	0.592	0.040 (0.047)	388
HH has latrine	0.293	-0.005 (0.042)	433	0.285	-0.026 (0.044)	388
HH floor is dirt	0.422	0.062 (0.046)	433	0.433	0.062 (0.048)	388
Joint F stat		0.651			0.740	
p-value		0.858			0.769	
Joint F stat (excl. T3ET)		0.823			0.947	
p-value		0.660			0.515	

Notes: Standard errors are reported in parentheses. Each coefficient estimate in Columns 2 and 4 comes from a regression of the covariates listed on the left-hand column on treatment and strata dummies. Columns 1 and 2 show results for the full sample. An F-test of joint significance is provided at the bottom of the table. This test includes only those students who have non-missing values for all variables. The bottom two rows report a joint test of significance that excludes T3ET scores. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 10: Covariate balance on treatment - P_{brig} sample

	Full Sample			Observed at Endline		
	Mean (1)	Diff. (2)	N (3)	Mean (4)	Diff. (5)	N (6)
T3ET English Score (Pct)	0.732	0.010 (0.009)	1,363	0.738	0.007 (0.010)	1,163
T3ET Math Score (Pct)	0.717	0.005 (0.010)	1,363	0.721	0.007 (0.010)	1,163
Female	0.485	0.000 (0.027)	1,413	0.482	0.002 (0.030)	1,204
Student has disability	0.007	0.003 (0.005)	1,413	0.006	0.005 (0.005)	1,204
Has sibling in Bridge	0.394	0.007 (0.025)	1,413	0.398	0.008 (0.027)	1,204
Mother is deceased	0.042	0.009 (0.010)	1,404	0.045	0.005 (0.011)	1,197
Father is deceased	0.164	-0.020 (0.016)	1,404	0.167	-0.025 (0.018)	1,197
<i>Caregiver Employment</i>						
Unemployed	0.216	0.043** (0.021)	1,413	0.210	0.061*** (0.023)	1,204
Worked in last 7 days	0.638	-0.034 (0.025)	1,413	0.645	-0.051* (0.027)	1,204
<i>Caregiver Occupation</i>						
Small business	0.189	-0.049*** (0.019)	1,413	0.196	-0.058*** (0.021)	1,204
Agriculture	0.127	-0.006 (0.016)	1,413	0.136	-0.004 (0.018)	1,204
Casual labor	0.375	-0.015 (0.024)	1,413	0.360	-0.023 (0.026)	1,204
<i>Other household characteristics</i>						
HH Income (000s Ksh)	53.322	0.019 (2.967)	1,413	54.169	-0.623 (3.325)	1,204
Caregiver can read	0.923	-0.028* (0.015)	1,413	0.923	-0.034** (0.016)	1,204
HH owns home	0.585	-0.020 (0.022)	1,404	0.563	-0.029 (0.025)	1,197
HH has electricity	0.674	-0.026 (0.023)	1,413	0.689	-0.027 (0.025)	1,204
HH has latrine	0.303	-0.003 (0.023)	1,413	0.315	-0.015 (0.026)	1,204
HH floor is dirt	0.518	-0.024 (0.024)	1,413	0.536	-0.025 (0.027)	1,204
Joint F stat		1.266			1.434	
p-value		0.202			0.107	

Notes: Standard errors are reported in parentheses. Each coefficient estimate in Columns 2 and 4 comes from a regression of the covariates listed on the left-hand column on treatment and strata dummies. Columns 1 and 2 show results for the full sample. An F-test of joint significance is provided at the bottom of the table. This test includes only those students who have non-missing values for all variables. The bottom two rows report a joint test of significance that excludes T3ET scores. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 11: Covariate balance on treatment - P_{priv} sample

	Full Sample			Observed at Endline		
	Mean (1)	Diff. (2)	N (3)	Mean (4)	Diff. (5)	N (6)
T3ET English Score (Pct)	0.760	0.018 (0.011)	1,042	0.762	0.022 (0.014)	766
T3ET Math Score (Pct)	0.781	-0.006 (0.011)	1,044	0.781	-0.001 (0.014)	768
Female	0.498	0.033 (0.029)	1,397	0.491	0.025 (0.034)	1,044
Student has disability	0.013	-0.011** (0.005)	1,397	0.010	-0.005 (0.006)	1,044
Has sibling in Bridge	0.083	0.029* (0.016)	1,397	0.102	0.017 (0.021)	1,044
Mother is deceased	0.090	-0.025* (0.013)	1,203	0.074	-0.023 (0.015)	891
Father is deceased	0.239	-0.010 (0.020)	1,203	0.223	-0.004 (0.024)	891
<i>Caregiver Employment</i>						
Unemployed	0.257	-0.021 (0.022)	1,397	0.261	-0.025 (0.026)	1,044
Worked in last 7 days	0.569	0.053** (0.025)	1,397	0.578	0.041 (0.030)	1,044
<i>Caregiver Occupation</i>						
Small business	0.170	0.029 (0.020)	1,397	0.169	0.020 (0.023)	1,044
Agriculture	0.073	-0.010 (0.013)	1,397	0.081	-0.010 (0.016)	1,044
Casual labor	0.379	-0.005 (0.025)	1,397	0.377	0.004 (0.030)	1,044
<i>Other household characteristics</i>						
HH Income (000s Ksh)	48.357	2.402 (2.770)	1,397	50.057	-1.194 (3.152)	1,044
Caregiver can read	0.913	-0.001 (0.015)	1,397	0.921	-0.002 (0.018)	1,044
HH owns home	0.697	0.020 (0.022)	1,203	0.683	0.004 (0.027)	891
HH has electricity	0.529	0.003 (0.024)	1,397	0.544	0.004 (0.029)	1,044
HH has latrine	0.287	-0.023 (0.022)	1,397	0.265	-0.017 (0.025)	1,044
HH floor is dirt	0.430	0.037 (0.025)	1,397	0.448	0.037 (0.029)	1,044
Joint F stat		1.611			1.178	
p-value		0.052			0.275	

Notes: Standard errors are reported in parentheses. Each coefficient estimate in Columns 2 and 4 comes from a regression of the covariates listed on the left-hand column on treatment and strata dummies. Columns 1 and 2 show results for the full sample. An F-test of joint significance is provided at the bottom of the table. This test includes only those students who have non-missing values for all variables. The bottom two rows report a joint test of significance that excludes T3ET scores. ***, **, and * indicate significance at 1%, 5%, and 10%.

7 Stage One Analysis

This section presents interim results that have been examined by the research team in preparing this document and that will guide the plan for the next stage of the analysis.

The effect of the scholarship on enrollment in Bridge and choices regarding other types of schools are examined first. We confirm that scholarship recipients are more likely to attend Bridge than non-recipients. However, in the absence of the scholarship, it appears that students might have attended a range of different education options, including government schools, non-Bridge private schools, and non-enrollment.

A large number of pupils attended Bridge even when they did not receive the scholarship. Because these students enrolled in Bridge, but also paid for Bridge, to consider the effect of the scholarship on *paying for Bridge* in addition to attendance at Bridge. For students who would have attended Bridge without the scholarship, the scholarship represents an infra-marginal transfer of income to the household.

This section will also present evidence on spillover effects of the scholarship on siblings of pupils who receive the scholarship. The results indicate that the scholarship increases the likelihood that both recipients and non-recipients attend Bridge.

Results are also presented on the effect of the scholarship extension in January 2018 on attendance at Bridge at the time of the interview. We find that students were likely to drop out when the scholarship program ended in January 2018 and that the extension induced many students to remain in Bridge.

Finally, this section will examine the effect of the scholarship on grade progression. Scholarship recipients appear to advance through grades at a more rapid pace. Scholarship recipients seem to be less likely to be held back.

7.1 The effects of the scholarship on educational environment

This section presents results on the first stage effect of the scholarship on the educational environment. Results are reported separately for samples PP_{govt} , P_{govt} , P_{brig} , P_{priv} , and P_{priv} .²⁷ Table 12 presents results on the effect of the scholarship on educational environment. Table 13 shows education environment by treatment and control. The first three columns of Table 12 illustrate the differences in educational choices of scholarship winners relative to non-winners for a pooled sample of PP_{govt} and P_{govt} pupils and PP_{govt} and P_{govt} separately. In the PP_{govt} and P_{govt} samples, the scholarship increased the probability that the pupil would enroll in Bridge by 36.8 percentage points. This effect is similar in both the PP_{govt} sample (36.3) and the P_{govt} sample (37.3).

²⁷The prefix PP indicates pre-primary, while P indicates primary school aged. The subscripts *govt*, *brig*, and *priv* correspond to samples that were more likely to attend government, Bridge, or non-Bridge private schools respectively. See Section 3.3.1 additional information.

Table 12: The effect of winning the scholarship on educational environment

	PP _{govt} +P _{govt} (1)	PP _{govt} (2)	P _{govt} (3)	PP _{brig} +P _{brig} (4)	PP _{brig} (5)	P _{brig} (6)	P _{priv} (7)
<i>Panel A: Attendance in 2016</i>							
Bridge	0.368 (0.010)	0.363 (0.015)	0.373 (0.013)	0.120 (0.017)	0.070 (0.033)	0.135 (0.019)	0.229 (0.032)
Government	-0.234 (0.009)	-0.117 (0.012)	-0.333 (0.014)	-0.075 (0.014)	-0.017 (0.025)	-0.093 (0.017)	-0.095 (0.027)
Private	-0.112 (0.007)	-0.199 (0.014)	-0.039 (0.008)	-0.045 (0.010)	-0.053 (0.025)	-0.042 (0.011)	-0.131 (0.028)
Unenrolled	-0.022 (0.003)	-0.047 (0.007)	0.000 (0.001)				-0.003 (0.005)
<i>Panel B: Attendance in 2017</i>							
Bridge	0.346 (0.010)	0.340 (0.015)	0.352 (0.013)	0.228 (0.021)	0.180 (0.042)	0.244 (0.024)	0.278 (0.031)
Government	-0.237 (0.010)	-0.144 (0.014)	-0.316 (0.014)	-0.160 (0.019)	-0.094 (0.033)	-0.182 (0.022)	-0.128 (0.029)
Private	-0.105 (0.008)	-0.188 (0.014)	-0.035 (0.008)	-0.068 (0.013)	-0.086 (0.032)	-0.062 (0.014)	-0.147 (0.027)
Unenrolled	-0.004 (0.002)	-0.007 (0.004)	-0.001 (0.002)				-0.003 (0.005)
<i>Panel C: Years of exposure</i>							
Bridge	0.714 (0.019)	0.702 (0.028)	0.725 (0.025)	0.348 (0.034)	0.249 (0.068)	0.379 (0.039)	0.507 (0.059)
Government	-0.471 (0.018)	-0.262 (0.024)	-0.649 (0.026)	-0.235 (0.030)	-0.111 (0.052)	-0.275 (0.035)	-0.223 (0.052)
Private	-0.218 (0.014)	-0.387 (0.025)	-0.074 (0.014)	-0.113 (0.022)	-0.139 (0.053)	-0.105 (0.023)	-0.277 (0.051)
Unenrolled	-0.026 (0.005)	-0.054 (0.009)	-0.002 (0.003)				-0.006 (0.010)
N	8,373	3,789	4,584	1,592	388	1,204	989

Notes: Each row represents a separate specification. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *govt*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. All specifications include strata fixed effects. Robust standard errors are reported in parentheses.

Table 13: Mean enrollment type for scholarship recipients and non-recipients

		PP _{gouv} +P _{gouv}	PP _{gouv}	P _{gouv}	PP _{brig} +P _{brig}	PP _{brig}	P _{brig}	P _{priv}
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A: Attendance in 2016</i>								
Bridge	Non-scholarship	0.256	0.331	0.197	0.808	0.841	0.797	0.392
	Scholarship	0.648	0.702	0.601	0.930	0.906	0.938	0.657
Government	Non-scholarship	0.510	0.256	0.712	0.123	0.067	0.142	0.262
	Scholarship	0.244	0.127	0.344	0.048	0.056	0.046	0.164
Private	Non-scholarship	0.203	0.345	0.089	0.069	0.091	0.061	0.338
	Scholarship	0.096	0.148	0.052	0.021	0.039	0.016	0.175
Unenrolled	Non-scholarship	0.031	0.068	0.002	0.000	0.000	0.000	0.008
	Scholarship	0.012	0.023	0.002	0.000	0.000	0.000	0.004
<i>Panel B: Attendance in 2017</i>								
Bridge	Non-scholarship	0.216	0.278	0.166	0.630	0.673	0.616	0.308
	Scholarship	0.583	0.629	0.544	0.857	0.839	0.863	0.620
Government	Non-scholarship	0.562	0.347	0.734	0.255	0.173	0.281	0.352
	Scholarship	0.297	0.188	0.390	0.100	0.089	0.104	0.222
Private	Non-scholarship	0.209	0.355	0.093	0.115	0.154	0.102	0.335
	Scholarship	0.109	0.168	0.059	0.043	0.072	0.033	0.153
Unenrolled	Non-scholarship	0.013	0.020	0.006	0.000	0.000	0.000	0.006
	Scholarship	0.010	0.015	0.006	0.000	0.000	0.000	0.004
<i>Panel C: Years of exposure</i>								
Bridge	Non-scholarship	0.472	0.609	0.363	1.438	1.514	1.414	0.700
	Scholarship	1.231	1.331	1.146	1.787	1.744	1.801	1.276
Government	Non-scholarship	1.072	0.604	1.446	0.378	0.240	0.423	0.614
	Scholarship	0.542	0.315	0.735	0.148	0.144	0.150	0.387
Private	Non-scholarship	0.412	0.700	0.182	0.184	0.245	0.164	0.673
	Scholarship	0.206	0.316	0.111	0.064	0.111	0.049	0.328
Unenrolled	Non-scholarship	0.044	0.088	0.009	0.000	0.000	0.000	0.013
	Scholarship	0.022	0.037	0.008	0.000	0.000	0.000	0.009
N		8,373	3,789	4,584	1,592	388	1,204	989

Notes: Table shows the share of students choosing each educational environment. Grade indicates the grade in which the student planned to enter at the start of the scholarship program. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *gouv*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship.

The effect of the scholarship on Bridge attendance represents pupils switching from government and non-Bridge private schools, as well as non-enrollment. In the pooled PP_{govt} and P_{govt} sample, the effect of the scholarship on Bridge attendance reflects students switching from both government and private schools where the first stage effect on attendance is -23.4 and -11.2 percentage points respectively. The effect of the scholarship on being unenrolled is -2.2 percent, suggesting that a few of students were induced to enroll in school by the scholarship.

There are important differences in these first stage responses to treatment across grades.²⁸ In the P_{govt} sample, the first stage effect on Bridge attendance (33.1 percentage points) is accounted for almost entirely by students switching out of government schools (-33.3). Moreover, in the P_{govt} sample, only 0.2 percent of pupils are unenrolled in 2016 and there is no discernible effect on non-enrollment. The estimated effect on the P_{govt} sample is therefore likely to closely approximate the relative effectiveness of government and bridge schools. In the PP_{govt} sample, the effect of the scholarship on Bridge attendance is accounted for predominantly by students induced to switch out of non-Bridge private schools (-19.9 percentage points) and the effect of the scholarship on being enrolled at all is 4.7 percentage points. The counterfactual for the PP_{govt} sample is therefore more mixed than the P_{govt} sample.

Most of these results are fairly stable across the two years of the intervention. As such, the first stage effect of the scholarship on the number of years of attendance at each option is approximately two times the first stage effect in either 2016 or 2017. For the PP_{govt} sample, a notable exception is that the effect of the scholarship on enrollment nearly disappears and is insignificant in 2017. This reflects the fact that most non-enrollment is in the earliest age cohort (those who were planning to enroll in “Baby class” in 2016) in the study and represents delayed entry into school. In 2016, 6.8 percent of pupils in the PP_{govt} sample who did not receive a scholarship were unenrolled compared to 2.0 percent in 2017.

The P_{brig} and P_{brig} samples were already enrolled in Bridge at the time of the application

²⁸Grade is defined as the grade the student planned to enroll in 2016 at the time of application in December 2015.

for the scholarship program and were likely to continue enrolling in Bridge, even if they did not receive the scholarship. For the pooled P_{brig} and P_{brig} sample, the first stage effect of the scholarship on attendance at Bridge was 12.0 percentage points in the first year and grew to 22.8 percentage points in the second year of the program. In the younger P_{brig} sample, this first stage effect on Bridge attendance comes largely from scholarship winners being less likely to drop out of Bridge and enroll in other private schools (-5.3 percentage points). In the older P_{brig} sample, the effect of the scholarship on government school attendance was -9.3 percentage points, compared to -4.2 percentage points for non-Bridge private school attendance. These results indicate that the first stage effect on Bridge attendance comes predominantly from scholarship winners being less likely to drop out and enroll in government schools, but also represents some students dropping out and enrolling in non-Bridge private schools.

The effect of the scholarship on Bridge attendance grows in the second year of the program as many scholarship non-recipients dropped out of Bridge. 80.8 percent of non-recipients in the P_{brig} and P_{brig} samples were enrolled in Bridge in 2016, compared to only 63.0 percent in 2017.

In the P_{priv} sample (pupils who were likely to enroll in non-Bridge private school in the absence of the scholarship), scholarship winners are 22.9 percentage points more likely to attend Bridge in 2016 and 27.8 percentage points more likely in 2017. This effect is considerably smaller than the corresponding effect on the P_{govt} sample. Non-recipients in the P_{priv} sample attended Bridge 39.2 percent of the time, compared to 19.7 percent of P_{govt} non-recipients. The rate of attendance among scholarship recipients was similar (60.1 percent for P_{govt} and 65.7 percent for P_{priv}). The first stage effect on non-Bridge private school attendance is larger than the effect for the P_{govt} sample (-13.1 percentage points compared to -3.9 percentage points); however, a large portion of the P_{priv} first stage effect on attendance at Bridge schools is accounted for by a negative effect on government school attendance, as well (-9.5 percentage points).

7.2 The effect on *paying for Bridge*

Many pupils attended Bridge even in the absence of the scholarship. In the nomenclature of program evaluation, pupils who attend Bridge even when they do not receive the scholarship are referred to as “always takers”. For these pupils, the scholarship did not change the school at which they attended, but it reduced the tuition bill that the pupil’s household paid.

Table 13 shows that 25.6 percent of scholarship non-recipients in the pooled PP_{govt} and P_{brig} sample attended Bridge in 2016, 21.6 percent attended in 2017. Scholarship non-recipients spent on average 0.472 years enrolled in Bridge schools over the two years of the scholarship. The share of “always takers” is larger in the younger PP_{govt} sample where scholarship non-recipients spent 0.609 years in Bridge schools (compared to 0.363 years in the older P_{govt} sample).²⁹

As expected, pupils in the pooled P_{brig} and P_{priv} sample (pupils who were already enrolled in Bridge at the time of the scholarship program) were even more likely to attend Bridge in the absence of the scholarship. 80.8 percent of non-recipients in this sample attended Bridge without the scholarship.

As already discussed above, the P_{priv} sample had a larger share of non-recipients attend Bridge than the P_{govt} sample with 39.2 percent of non-recipients attending Bridge.

7.3 Sibling spillover effects on attendance

Many applicants live in households with other applicants. The scholarship may have had important spillover effects on students within the same household. This sub-section presents results on the effects of siblings on attendance at Bridge. To begin, the method for identifying siblings in the data is described. Second, an empirical test of the effect of sibling spillovers on attendance is described. Finally, results of this test are presented.

²⁹This result was expected. The P_{brig} sample included pupils who said that they would go to Bridge in the absence of the scholarship, whereas the P_{govt} sample excluded applicants who said they would go to Bridge in the absence of the scholarship.

While data on household membership was not collected at baseline, phone numbers and contact information can be used to identify children linked by the same adult caregivers. Prior to endline activities, households were identified using caregiver contact information for logistical purposes, and the procedure was found to be accurate at predicting household membership. To maximize the sample size used for the estimation of sibling spillovers, the procedure used to identify shared households is repeated using the full application data file.

These households can be used to test for spillovers within the household onto other scholarship participants. 71 percent of applicants in the study are found to be in a household with another applicant using this method. Over 90 percent of these households have 5 or fewer applicants. A few “households” are found to have more than 15 children. These households may be multi-family homes, polygamous homes, or orphanages.³⁰

To illustrate the effect spillovers on choice of education environment, consider the subsample of control students who are in multi-applicant households (two or more applicants). Let $h(i)$ be a function that identifies the household associated with applicant i . Denote the number of other applicants in applicant i 's household as $N_{h(i)}$, and the share of other household applicants treated as

$$\zeta_i = \frac{\sum_{k:h(i)=h(k),i \neq j} Z_i}{N_{h(i)} - 1}$$

Siblings in different strata may have different propensities to receive treatment. To account for this, and to ensure that measure of the degree to which siblings in the household receive scholarships is conditionally independent of potential outcomes, transform ζ_i into a binary variable indicating that it is above some threshold and form a corresponding propensity score using simulation of the scholarship assignment randomization procedure. Defining ζ_i^{25} to be an indicator for whether more than 25 percent of pupil i 's siblings received the

³⁰In some cases, large numbers of applications using the same phone number may be indications of “phantom” applications where, for example, someone might submit several applications and in the hopes of winning some and reselling scholarships later. However, if this were the case, we would expect that the follow-up rate would be much lower among groups of children using the same phone number. However, there is not a clear relationship between follow-up rates and the number of applications using the same phone number.

scholarship, a propensity score control can be obtained by simulating the randomization procedure described in Section 3.3 S times and then estimating the frequency with which ζ_i exceeds 0.25.

The effect on i of offering a scholarship to siblings in the household $k \neq i$ is estimated by

$$D_i = \alpha_0 + \alpha_1 \zeta_i^{25} + \alpha_2 P_i(\zeta_i^{25}) + \epsilon_i$$

Where $P_i(\zeta_i^{25})$ represents the propensity score and D_i is any endogenous attendance variable.³¹

³¹Note that this is one of many ways to use the experimental variation in treatment status of siblings for identification. Another simple approach is to define ζ_i as an indicator for whether *any* sibling is treated.

Table 14: The spillover effect of sibling scholarship (ξ^{25}) on educational environment for non-recipients

	PP _{govt} +P _{govt} (1)	PP _{govt} (2)	P _{govt} (3)	PP _{brig} +P _{brig} (4)	PP _{brig} (5)	P _{brig} (6)	P _{priv} (7)
N	2,740	1,203	1,537	1,076	618	921	339
<i>Panel A: Attendance in 2016</i>							
Bridge	0.115*** (0.021)	0.091*** (0.034)	0.131*** (0.025)	0.035 (0.027)	0.039 (0.034)	0.028 (0.029)	0.205*** (0.078)
Government	-0.077*** (0.023)	-0.030 (0.031)	-0.097*** (0.028)	-0.016 (0.022)	0.005 (0.026)	-0.013 (0.024)	-0.048 (0.072)
Private	-0.042*** (0.018)	-0.062* (0.033)	-0.036** (0.018)	-0.017 (0.018)	-0.040* (0.024)	-0.013 (0.019)	-0.134* (0.077)
Unenrolled	0.003 (0.008)	0.001 (0.019)	0.002 (0.003)				-0.023 (0.015)
<i>Panel B: Attendance in 2017</i>							
Bridge	0.102*** (0.020)	0.097*** (0.033)	0.104*** (0.023)	0.062* (0.033)	0.051 (0.042)	0.058 (0.036)	0.161** (0.074)
Government	-0.051** (0.023)	-0.029 (0.034)	-0.056** (0.027)	-0.029 (0.029)	-0.001 (0.036)	-0.024 (0.032)	-0.099 (0.078)
Private	-0.046*** (0.018)	-0.064* (0.033)	-0.041*** (0.018)	-0.031 (0.023)	-0.046 (0.030)	-0.032 (0.024)	-0.043 (0.077)
Unenrolled	-0.005 (0.005)	-0.004 (0.010)	-0.006 (0.005)				-0.019 (0.013)
<i>Panel C: Years of exposure</i>							
Bridge	0.217*** (0.037)	0.188*** (0.061)	0.235*** (0.044)	0.097* (0.054)	0.090 (0.069)	0.086 (0.059)	0.366*** (0.143)
Government	-0.128*** (0.043)	-0.059 (0.060)	-0.154*** (0.052)	-0.045 (0.046)	0.004 (0.056)	-0.036 (0.051)	-0.147 (0.142)
Private	-0.088*** (0.034)	-0.126** (0.061)	-0.077*** (0.032)	-0.048 (0.037)	-0.086* (0.049)	-0.044 (0.039)	-0.177 (0.148)
Unenrolled	-0.002 (0.011)	-0.003 (0.024)	-0.004 (0.007)				-0.042 (0.027)

Notes: Each row represents a separate specification. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *govt*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. All specifications include a linear treatment propensity score (in place of strata fixed effects) and a fixed effect for the number of pupils in the household. Specification restricts to households with between 2 and 5 applicant pupils. Robust standard errors are reported in parentheses. PP_{brig} and P_{brig} sample includes siblings of randomly sampled students. ***, **, and * indicate significance at 1%, 5%, and 10%.

This test indicates that siblings of scholarship recipients were much more likely to attend Bridge. Table 14 shows the results of the specification above for scholarship non-recipients. Among non-recipient applicants in the PP_{govt} and P_{govt} pooled sample, when at least 25 percent of siblings are treated ($\zeta_i^{25} = 1$), an applicant spent 0.217 more years in Bridge schools. This effect comes from non-recipient siblings switching out of government schools and non-Bridge private schools. For P_{govt} pupils, pupils spend 0.235 more years in Bridge when more than 25 percent of their siblings are assigned scholarships, and most of this effect comes from students moving from government schools (-0.154 for government schools, compared to -0.077 for private schools).

There are similar effects in the P_{brig} and P_{priv} samples, although the sample sizes are smaller. The small number of pupils in the P_{priv} sample who have siblings exhibit very large effects from assignment of scholarship on treatment. In this sample, pupils spend 0.366 additional years in Bridge schools when more than 25 percent of their siblings are treated.

Table 15 shows analogous results for the sample of scholarship recipients. Scholarship recipients are also more likely to attend Bridge when their siblings receive scholarships.

Table 15: The spillover effect of sibling scholarship (ζ^{25}) on educational environment for scholarship recipients

	PP _{govt} +P _{govt} (1)	PP _{govt} (2)	P _{govt} (3)	PP _{brig} +P _{brig} (4)	PP _{brig} (5)	P _{brig} (6)	P _{priv} (7)
N	3,183	1,405	1,778	745	354	627	306
<i>Panel A: Attendance in 2016</i>							
Bridge	0.047*** (0.020)	0.045 (0.030)	0.048* (0.028)	0.063*** (0.024)	0.051 (0.036)	0.057** (0.026)	0.138 (0.084)
Government	-0.031* (0.018)	0.008 (0.023)	-0.054** (0.027)	-0.047*** (0.020)	-0.041 (0.030)	-0.040* (0.022)	-0.157*** (0.066)
Private	-0.012 (0.012)	-0.047** (0.023)	0.008 (0.013)	-0.017 (0.014)	-0.010 (0.023)	-0.016 (0.014)	0.039 (0.067)
Unenrolled	-0.003 (0.004)	-0.005 (0.009)	-0.003 (0.002)				-0.020** (0.010)
<i>Panel B: Attendance in 2017</i>							
Bridge	0.084*** (0.021)	0.097*** (0.032)	0.075*** (0.028)	0.087*** (0.032)	0.121*** (0.046)	0.077** (0.035)	0.155* (0.086)
Government	-0.056*** (0.020)	-0.027 (0.026)	-0.074*** (0.027)	-0.081*** (0.028)	-0.111*** (0.038)	-0.065** (0.031)	-0.165** (0.075)
Private	-0.017 (0.013)	-0.054*** (0.023)	0.006 (0.013)	-0.007 (0.019)	-0.009 (0.031)	-0.012 (0.019)	0.030 (0.062)
Unenrolled	-0.011*** (0.004)	-0.016* (0.008)	-0.007* (0.004)				-0.020** (0.010)
<i>Panel C: Years of exposure</i>							
Bridge	0.130*** (0.040)	0.141*** (0.059)	0.123*** (0.054)	0.151*** (0.052)	0.172*** (0.076)	0.134*** (0.056)	0.294* (0.165)
Government	-0.088*** (0.036)	-0.019 (0.046)	-0.128*** (0.052)	-0.127*** (0.044)	-0.152*** (0.063)	-0.106** (0.048)	-0.322*** (0.132)
Private	-0.029 (0.023)	-0.101*** (0.042)	0.015 (0.024)	-0.023 (0.030)	-0.020 (0.050)	-0.028 (0.031)	0.069 (0.122)
Unenrolled	-0.014* (0.008)	-0.021 (0.016)	-0.010* (0.006)				-0.041** (0.020)

Notes: Each row represents a separate specification. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *govt*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. All specifications include a linear treatment propensity score and a fixed effect for the number of pupils in the household. Samples are restricted to households with between 2 and 5 applicant pupils. Robust standard errors are reported in parentheses. PP_{brig} and P_{brig} sample includes siblings of randomly sampled students. ***, **, and * indicate significance at 1%, 5%, and 10%.

7.4 Effect of scholarship extension on enrollment

Endline activities were conducted from late November 2017 to March 2018. Some former scholarship students were likely to be switching schools at that time as the term of the scholarship was ending. 35 percent of all endline surveys and 48 percent of all those for the PP_{govt} and P_{govt} pupils were conducted prior to the start of the 2018 academic calendar. Bridge extended 2,300 scholarships for pupils in the PP_{govt} and P_{govt} to continue attending Bridge schools for one term on scholarship. Pupils in the P_{govt} sample who were in the last two grades of primary school (standards 5 through 8) were all provided the extension. The remaining scholarships were randomly assigned at the academy level.³²

This section examines whether the scholarship had an effect on educational environment around the time of the interview. Four endogenous school attendance outcomes are considered:

- Attendance at Bridge in 2018
- An indicator for whether the pupil changed schools in 2018
- An indicator for whether they changed school types in 2018
- An indicator for whether the pupil left Bridge between 2017 and 2018

To begin, we will examine whether the scholarship had an effect on any of these variables for the sample of students who were interviewed in 2018. This analysis will show that pupils may have been more likely to switch schools or school types as a result of the original scholarship offer. However, these results also confirm that school switching is common among both treated and control pupils. The scope for school switching effects to influence the results are quite limited. After that we will examine whether the scholarship extension influenced these effects. This analysis will be composed of two parts. We will look at whether the scholarship extension reduced this effect. Because the extension was only randomized for

³²Randomization was stratified by province.

pupils in Standards 6 and below, we restrict this analysis to those who were projected to be in these grades according to baseline data.³³

Table 16 shows the results of estimating the effect of the scholarship on Bridge attendance in 2018 and indicators of school environment switching. Winning the scholarship increased the probability that a student in the PP_{govt} and P_{govt} pooled sample attended Bridge in 2018 by 22.6 percentage points, about two-thirds of the effect in earlier years. Scholarship winners were 10.1 percent more likely to leave Bridge in 2018. However, scholarship winners were only 6.3 percent more likely to change school types and 4.6 percent more likely to change schools. Changing schools between grades appears to be quite common. 27.8 percent of scholarship non-recipients changed schools in 2018, so the total effect of the scholarship on changing schools is small.

³³The extension itself was assigned according to an endogenous characteristic: grade in 2018. The projected grade in 2018 according to baseline application data is highly predictive of endline grade and is not potentially endogenous to treatment.

Table 16: Effect of the scholarship on school switching in 2018

	$PP_{govt}+P_{govt}$ (1)	PP_{govt} (2)	P_{govt} (3)	$PP_{brig}+P_{brig}$ (4)	PP_{brig} (5)	P_{brig} (6)	P_{priv} (7)
N	3,817	1,842	1,975	1,304	340	964	922
<i>Bridge school</i>							
Treat	0.194*** (0.030)	0.195*** (0.039)	0.193*** (0.044)	0.141*** (0.054)	0.102 (0.114)	0.152*** (0.061)	0.148** (0.070)
Treat × Extension	0.073** (0.036)	0.097** (0.047)	0.028 (0.054)	-0.007 (0.065)	0.049 (0.128)	-0.029 (0.078)	0.013 (0.082)
<i>Changed schools in 2018</i>							
Treat	0.043 (0.029)	0.009 (0.038)	0.093** (0.043)	0.071 (0.047)	0.035 (0.098)	0.081 (0.053)	-0.001 (0.067)
Treat × Extension	0.000 (0.035)	0.004 (0.045)	0.004 (0.053)	-0.013 (0.057)	-0.046 (0.110)	0.026 (0.068)	0.069 (0.079)
<i>Changed school types in 2018</i>							
Treat	0.050* (0.027)	0.040 (0.037)	0.064 (0.040)	0.039 (0.046)	-0.019 (0.092)	0.055 (0.054)	0.021 (0.066)
Treat × Extension	-0.040 (0.033)	-0.062 (0.044)	0.007 (0.050)	0.021 (0.056)	0.026 (0.103)	0.042 (0.069)	0.058 (0.077)
<i>Left Bridge in 2018</i>							
Treat	0.086*** (0.022)	0.097*** (0.029)	0.069** (0.034)	0.026 (0.043)	-0.095 (0.082)	0.059 (0.051)	0.065 (0.055)
Treat × Extension	-0.033 (0.027)	-0.069** (0.035)	0.031 (0.043)	0.071 (0.052)	0.142 (0.092)	0.073 (0.065)	0.057 (0.064)

Notes: The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *govt*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. All specifications include strata fixed effects. Sample is restricted to those who were interviewed after the start of the 2018 academic year. ***, **, and * indicate significance at 1%, 5%, and 10%.

The school-level randomized extension appears to have done little to reduce school switching. We use the following specification to test for an interaction effect between scholarship assignment and scholarship extension.

$$D_i = \lambda_s + \alpha Z_i + \beta Z_i \times E_j + \delta E_j + \epsilon_i$$

Where λ_s is a strata fixed effect and Z_i indicates receipt of the scholarship and E_j indicates that the pupil i 's school j was selected for the extension.

We restrict to the sample of students who were planning to enroll in Standard 4 or lower in 2016. This choice is motivated out of a concern that some switching of schools may arise from progress into secondary school.³⁴ These results also restrict to pupils who were interviewed *after* January 5, 2018, the beginning of the extension period. Pupils were notified of the scholarship extension starting on this date.

³⁴The results are largely unchanged when leaving the older pupils in the sample.

Table 17: The effect of the scholarship and extension on 2018 school switching

	$PP_{gout}+P_{gout}$ (1)	PP_{gout} (2)	P_{gout} (3)	$PP_{brig}+P_{brig}$ (4)	PP_{brig} (5)	P_{brig} (6)	P_{priv} (7)
N	3,817	2,755	1,842	913	1,062	1,304	340
<i>Panel A: Bridge school</i>							
Treat	0.173*** (0.025)	0.192*** (0.031)	0.187*** (0.040)	0.203*** (0.047)	0.118*** (0.045)	0.198*** (0.045)	0.089 (0.114)
Treat × Extension	0.077*** (0.031)	0.071* (0.037)	0.105** (0.047)	-0.006 (0.059)	0.098* (0.054)	-0.078 (0.055)	0.069 (0.127)
<i>Panel B: Government school</i>							
Treat	-0.133*** (0.028)	-0.121*** (0.033)	-0.058 (0.041)	-0.232*** (0.055)	-0.169*** (0.054)	-0.170*** (0.043)	-0.160* (0.095)
Treat × Extension	-0.032 (0.034)	-0.026 (0.040)	-0.093* (0.049)	0.094 (0.069)	-0.045 (0.064)	0.122*** (0.052)	0.106 (0.106)
<i>Panel C: Non-Bridge Private</i>							
Treat	-0.047** (0.023)	-0.072*** (0.029)	-0.114*** (0.039)	0.001 (0.038)	0.027 (0.034)	-0.023 (0.031)	0.103 (0.084)
Treat × Extension	-0.032 (0.028)	-0.035 (0.035)	-0.018 (0.046)	-0.048 (0.048)	-0.033 (0.041)	-0.047 (0.038)	-0.215*** (0.094)
<i>Panel D: Unenrolled</i>							
Treat	0.007 (0.012)	0.001 (0.014)	-0.015 (0.019)	0.028* (0.017)	0.024 (0.026)	-0.005 (0.007)	-0.032 (0.020)
Treat × Extension	-0.013 (0.015)	-0.011 (0.017)	0.006 (0.023)	-0.040* (0.021)	-0.020 (0.031)	0.003 (0.008)	0.041* (0.022)

Notes: Each panel represents a separate specification. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *gout*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. All specifications include strata fixed effects. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 17 presents results of this analysis. The results show that, in the PP_{govt} and P_{govt} pooled sample, scholarship recipients at non-extension schools were 19.4 percentage points more likely to be attending Bridge than non-recipients. At extension schools, this figure was 7.3 percentage points higher. There are similar effects of the treatment on 2018 attendance for the P_{brig} , P_{brig} and P_{priv} samples, but, in these samples, there is not evidence that the scholarship extension affected the relationship between initial scholarship receipt and Bridge attendance.

Relatedly, the scholarship may have increased likelihood that pupils switched schools in 2018. In the pooled PP_{govt} and P_{govt} sample, pupils were 4.3 percent more likely to switch schools in 2018, although this effect is not statistically significant. There does not appear to be any effect of the extension on this outcome. In the P_{govt} sample, treatment induced 9.3 of students to switch schools in 2018.

PP_{govt} and P_{govt} pupils who received the scholarship were 8.6 percentage points more likely to leave Bridge in 2018 and the extension only appears to have had a small (and insignificant) -3.3 percentage point effect on retaining students. This effect is larger in the PP_{govt} sample, where the scholarship increased the likelihood that a pupil would be leaving Bridge in 2018 by 9.7 percentage points. For the PP_{govt} sample, the scholarship extension appears to have mitigated this effect: at extension schools this effect is 6.9 percentage points lower.

7.5 Effects of treatment on grade progression

This section presents results of the effects of the scholarship on grade progression. Analysis of phone call data from 2016 and reports from the field team indicated that many pupils were enrolled in grades that differed from those that would have been expected given the data provided at baseline, and receipt of the scholarship seemed to induce students to make more rapid progress through grades.

Using baseline application data, it is possible to project the grade in which each pupil

would have been found at endline if they had (a) enrolled in the grade which they had planned to in 2016 and (b) had passed that grade and successfully enrolled in the next grade in 2017. Using this projection, we partition the sample according to the grade in which we would have expected the pupil to be in 2017, at endline. For each projected grade, we examine the share of pupils who were in each grade level at endline.

The results presented in this section are intended to shed light on how the scholarship influenced the grade level of the assessment that the pupil sat for at endline. These results have been analyzed beforehand in order to understand potential risks to identification. In the results that follow, ECD is treated as a single grade because all ECD pupils received the same test. Subsequent analysis will expand on this analysis by distinguishing between different grade levels within the ECD category.

Table 18 presents these descriptive statistics. While most pupils reach their projected grade, many do not. The remainder of this section will examine whether treatment influenced grade progression and, as a result, the assessment level the pupil received at endline.

Table 18: Grade progression matrix

Projected Grade	Grade achieved/assessment level									N (10)
	ECD (1)	STD 1 (2)	STD 2 (3)	STD 3 (4)	STD 4 (5)	STD 5 (6)	STD 6 (7)	STD 7 (8)	STD 8 (9)	
ECD	0.920	0.064	0.009	0.002	0.002	0.001	0.000	0.002	0.001	2,909
STD 1	0.277	0.631	0.078	0.005	0.003	0.001	0.003	0.000	0.001	880
STD 2	0.062	0.194	0.704	0.038	0.000	0.002	0.000	0.000	0.000	625
STD 3	0.010	0.041	0.173	0.696	0.066	0.004	0.003	0.006	0.001	724
STD 4	0.006	0.003	0.022	0.134	0.788	0.037	0.007	0.001	0.001	722
STD 5	0.005	0.003	0.002	0.033	0.150	0.750	0.050	0.008	0.000	639
STD 6	0.002	0.000	0.000	0.008	0.020	0.162	0.767	0.032	0.011	665
STD 7	0.003	0.003	0.000	0.003	0.008	0.026	0.215	0.698	0.043	605
STD 8	0.000	0.002	0.002	0.003	0.005	0.002	0.043	0.242	0.702	604

Notes: Projected grade is obtained from baseline data that the pupil submitted at the application stage. If an applicant planned to attend Standard 1 in 2016, then their projected grade at endline is Standard 2. Pupils in Baby Class and Nursery Class are projected to be in the final year of pre-primary (Preunit) and would sit for the ECD test. Grade achieve/assessment level refers to the grade that the pupil was found in at endline. All students in Baby Class, Nursery Class, or Preunit took the ECD form of the assessment.

We begin by estimating the effect of the scholarship on the grade the pupil is in at endline. We estimate the following reduced form model:

$$A_i = \alpha Z_i + X_i' \beta + \epsilon_i \quad (1)$$

Where A_i is the integer class level attended by pupil i at endline, $Z_i \in \{0, 1\}$ indicates scholarship status, X_i is a $k \times 1$ column vector of baseline covariates, including a dummy variable indicating the class the pupil intended to attend in 2016, and ϵ_i is a disturbance term.

Table 19 shows the results of this analysis in each of the samples discussed above. The results indicate that scholarship winners are significantly more likely to sit for a more advanced assessment than scholarship non-recipients. This effect appears to be concentrated in the P_{govt} sample.

Table 19: Effect of treatment on grade progression

Sample	Full Sample			High contact sample			IPA recruited		
	Share projection accurate	Coef.	N	Share projection accurate	Coef.	N	Share projection accurate	Coef.	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PP _{govt}	0.844	-0.007 (0.019)	3,789	0.855	-0.028 (0.021)	2,717	0.815	0.036 (0.038)	713
P _{govt}	0.684	0.105*** (0.023)	4,584	0.707	0.090*** (0.027)	3,155	0.692	0.132*** (0.040)	973
PP _{govt} + P _{govt}	0.755	0.054*** (0.015)	8,373	0.774	0.034* (0.017)	5,872	0.742	0.090*** (0.028)	1,686
PP _{brig}	0.846	-0.006 (0.038)	388	0.846	-0.006 (0.038)	388	1.000	-	7
P _{brig}	0.896	0.031 (0.020)	1,204	0.896	0.031 (0.020)	1,204	1.000	-	17
PP _{brig} + P _{brig}	0.884	0.022 (0.018)	1,592	0.884	0.022 (0.018)	1,592	1.000	-	24
P _{priv}	0.757	0.005 (0.048)	989	0.747	0.028 (0.056)	671	0.716	0.226* (0.122)	157

Notes: This table presents results from specifications following Equation 1. The table shows results for sub-samples of the experiment. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscripts *govt*, *brig*, and *priv* refer to pupils who were determined, using baseline data from the application, to be likely to attend government, Bridge, or non-Bridge private schools (respectively) in the absence of a scholarship. The Full sample refers to the entire sample of students to be included in the analysis. This study will also explore results with a “High contact” sample of students who were observed either in Bridge data or interacted with IPA staff prior to scholarship assignment. The IPA recruited sample refers to just those applicants who were applied to the study through IPA.

To better understand this effect, we test the effect of treatment on indicators for each grade level using multivariate regression. Using baseline data on the grade in which each pupil planned to enroll, we identify the “projected grade” for each pupil. We then split the sample according to this projected grade, and then estimate the effect of treatment on each endline grade achievement indicator in a multivariate regression model.

The results of this analysis are presented in Table 20. The results indicate that pupils are more likely to be in their projected grade if they receive the scholarship. Most of this effect appears to come from pupils being made less likely to be in lower grades, although there is some evidence that the scholarship may also make students less likely to be in higher grades as well.

Table 20: Grade Progression Effect Matrix- Multivariate regression results - PP_{gout} and P_{gout} Sample

Projected assessment	Grade achieved/Assessment taken									
	ECD (1)	STD 1 (2)	STD 2 (3)	STD 3 (4)	STD 4 (5)	STD 5 (6)	STD 6 (7)	STD 7 (8)	STD 8 (9)	N (10)
STD 1	-0.032 (0.031)	0.056* (0.034)	-0.016 (0.019)	-0.001 (0.005)	-0.004 (0.004)	-0.002 (0.002)	0.001 (0.004)	-	-0.003 (0.002)	880
STD 2	-0.034 (0.021)	-0.072** (0.033)	0.139*** (0.038)	-0.031* (0.016)	-	-0.003 (0.003)	-	-	-	625
STD 3	-0.011 (0.008)	-0.029* (0.016)	-0.050* (0.030)	0.050 (0.036)	0.034* (0.019)	-0.003 (0.005)	0.001 (0.004)	0.005 (0.006)	0.002 (0.003)	724
STD 4	-0.007 (0.006)	-0.004 (0.004)	-0.027*** (0.011)	-0.040 (0.027)	0.097*** (0.032)	-0.012 (0.015)	-0.008 (0.006)	-0.003 (0.003)	0.004 (0.003)	722
STD 5	-0.012** (0.005)	0.000 (0.005)	0.003 (0.003)	-0.002 (0.014)	-0.086*** (0.029)	0.126*** (0.035)	-0.026 (0.018)	-0.003 (0.007)	-	639
STD 6	0.003 (0.003)	-	-	-0.008 (0.007)	-0.025** (0.011)	-0.029 (0.030)	0.067* (0.035)	-0.004 (0.014)	-0.005 (0.008)	665
STD 7	0.001 (0.005)	0.006 (0.005)	-	-0.003 (0.005)	-0.001 (0.008)	-0.031*** (0.014)	0.004 (0.035)	0.038 (0.040)	-0.013 (0.017)	605
STD 8	-	-0.005 (0.003)	-0.003 (0.003)	-0.007 (0.005)	0.003 (0.006)	-0.003 (0.003)	0.005 (0.017)	-0.033 (0.037)	0.043 (0.039)	604

Notes: Each row represents a separate multivariate regression specification. The row labels indicate the sample used in the analysis. The column labels indicate the dependent variables. Projected assessment levels for pupils are calculated using baseline data on the pupil's academic plans.

8 Intent of Plan for Subsequent Analysis

The second stage of the analysis was planned while maintaining constraints on data access. The intent of this process is to ensure that decisions regarding data sources, cleaning, and methods for subsequent analysis are made without knowing the influence of those decisions on estimated treatment effects on outcomes to be analyzed in the next stage.

The plan will describe hypotheses that will be tested. The data collected at endline address diverse concerns of many stakeholders, each of which may have different social preferences and prior beliefs. Stakeholders will have their own decision criteria for whether and how Bridge should be regulated, prohibited, and/or financed. Their views on appropriate policy may well depend on their views on many different issues beyond those which a causal analysis of the impact of winning a scholarship to Bridge can address, such as the appropriate role of parental choice in education and the labor practices of Bridge. Stakeholders' judgments about how and whether to regulate, prohibit, or finance Bridge may also depend on their views about the relative importance of a range of educational outcomes such as scores on various different subject tests and grade progression, as well as how they weigh effects across students with different initial test scores, genders, and other characteristics. An analysis of whether effects on a single primary outcome pre-specified by researchers are significantly different from zero at a 5 percent significance level is unlikely to be an appropriate criterion for policy decisions regarding Bridge.

In our view, stakeholders should use the results of this study to update priors based on their own views about the trade-off between different types of outcomes and the totality of the data, not based on pre-analysis choices made by researchers about what to consider as the "primary outcome." Thus, for example, policy makers may weigh different subjects, such as English, Kiswahili, and Math, based on their own views of the importance of each subject, rather than based on the weights that researchers happened to pre-specify. Similarly, to the extent that different policy makers and analysts have different priors on the

processes determining attrition, they will focus on different approaches for dealing with attrition. Pre-specifying all the possible outcome measures, controls, and procedures for dealing with statistical issues such as differential attrition and grade repetition is, as Olken (2015) discusses, nearly impossible.

There is a tension between pre-specifying an analysis plan and the previously standard approach in the economics profession in which researchers conduct a range of robustness tests in their first draft, present preliminary results at seminars, conduct additional follow-up tests based on seminar comments, and then conduct additional tests in response to comments from referees and editors (Olken 2015). The standard approach in economics involves running multiple specifications to give the reader a sense of robustness, and to shed light on mechanisms. Under a pre-analysis plan reporting a range of outcome measures would require multiple hypothesis testing procedures, which would reduce reported power (Anderson and Magruder, 2017). We think this is misleading when specifications are to test robustness and shed light on mechanisms.

Because of concerns regarding both the standard approach in economics and the pre-analysis plan approach, we have opted for a third, sequential analysis, approach in which we first analyze certain outcomes while keeping ourselves blinded to others and outline a plan for subsequent analysis, and then conduct the subsequent analysis. This will constrain our ability to select estimates in a way that leads to smaller or larger estimated test score effects. Because we will nonetheless report multiple outcomes and specifications, as in the standard approach in economics, we are likely to reject at least one null hypothesis with $p > 0.05$ even if all null hypotheses are true.

8.1 Research team data access

In order to more effectively develop this first stage analysis and plan for subsequent analysis, the research team had limited and proscribed access to applicant microdata, including endline test scores. This has allowed us to better understand the first stage response to treatment

and determine in advance what covariates may be helpful to improve the statistical precision of endline effect estimates; to examine and develop procedures for dealing with differential grade progression across treatment arms; to better characterize the counterfactual schooling choices against which Bridge is being compared and elaborate plans to decompose the overall treatment effect to estimate these separate effects; and to explore whether and to what extent attrition was a problem and develop plans to address it as well. Throughout, the research team has remained blind to treatment when examining outcome data like test scores and other endline data described in Section 10.³⁵

Specifically, the data have been analyzed from two phone call surveys with study participants. At the end of each school year, a phone call tracking survey was conducted with applicants to the scholarship program. These phone call surveys were primarily intended to maintain contact with applicants to update contact information and improve follow-up at endline. The phone call surveys also collected data on pupil school attendance. These data have been used to examine the first stage effect of the scholarship on attendance at a Bridge school. To maximize the statistical precision of the endline effect estimates, the research team has tested whether baseline covariates can be used to improve the first stage response of treatment.

In addition, the research team has also used this data to understand the counterfactual against which Bridge is being compared. The phone call survey asked respondents whether they were attending school at all, and if so whether the school is a government school, a Bridge school, or a non-Bridge private school. Based on these data, the research team has already evaluated the effect of treatment on the decision to attend each of these mutually exclusive school types. This analysis was critical to developing plans to be able to decompose the effect of treatment on each of the counterfactual outcomes.

The phone call data have also been used to examine differential attrition and non-consent rates by treatment. Prior to the phone call survey, there was a concern that many potential

³⁵As will be discussed at length below, the research team has examined the effect of treatment on school attendance, grade progression, and an indicator variable for endline follow-up.

applicants may be difficult to track over a long period in a study that spanned the entire country. Therefore, it was important to carefully monitor the aggregate follow-up rate and test whether the treatment had a differential impact on follow-up rates.

Finally, the research team has examined the correlation between endline test score outcomes and baseline data to select controls to include in the main analysis. This analysis was conducted using endline data from the first approximately 30 percent of applicants surveyed in the endline. Using this limited portion of endline data, the research team examined the correlation between baseline data from the application phase of the scholarship program and endline test scores in mathematics and English. This activity examined scores of both treatment and control pupils due to the small sample size available at the time, but the research team remained blind to treatment status throughout.

9 Empirical Framework for Estimation of Causal Effects of the Scholarship

We describe the plan of analysis for estimation of the causal effect on test scores because this outcome is likely to be salient in public discussions. Advance planning for the analysis is intended as a precaution against the potential influence of subconscious bias on the selection of estimators for this effect that could make the scholarship program appear to have a larger or smaller impact on test scores. The possibility of bias in choice of estimators regarding the effect of the scholarship on education environment, for example, seems smaller.

This section describes the plan of analysis for the effect of expanding household budget sets by means of a scholarship and encouraging pupils to newly enroll in a Bridge school. This effect is informative about a broad range of policy choices (e.g. vouchers) that would operate by changing the costs of attending Bridge for families. After discussing the identification of this effect, the interpretation of the effect will be discussed in depth and extensions will be considered. Reducing the costs of Bridge relative to other options induces many students

to attend Bridge, but the interpretation of an observed effect is complicated by the fact that students in the control group were free to choose any educational option. Control students primarily attended government schools, but many attended private schools and some were unenrolled. This section will describe approaches that will be used to decompose the scholarship effect to make comparisons between specific schooling alternatives and Bridge, e.g. between government primary schools and Bridge primary schools. Finally, controls to be included across all specifications will be discussed.

9.1 The Scholarship Effect

In overall design, this study is closely related to other RCTs that provide a scholarship to lower costs to induce a particular behavior (Rouse (1998); Witte (1998); Abdulkadiroğlu et al. (2018); Dean and Jayachandran (2019)). Because the scholarships covered fees at Bridge schools, and thus lowered the cost of attending, the lottery creates variation in Bridge attendance across winners and non-winners. This experimental design allows for clear identification of the effect of a policy of providing scholarships that expand the budget set by reducing the cost of some private schools. It is common, in the study of scholarship programs, to assume that the scholarship itself does not influence outcomes beyond its effect on attendance. When this can be assumed, it is possible to use 2SLS to then identify the effect of attending Bridge on those students for whom the scholarship changed their choice of school.

A comparison of student outcomes across lottery winners and non-winners yields a measure of the effects of the scholarship program itself. This is an Intention-To-Treat (ITT) effect and can be estimated with the following linear model:³⁶

$$Y_i = \delta_0 + \delta_1 Z_i + \epsilon_i \tag{2}$$

³⁶For ease of exposition, the following section ignores additional baseline covariate and stratification controls. These will be discussed in section 9.5.

where $Z_i \in \{0, 1\}$ indicates a student’s treatment status. The estimate of δ_1 , $\hat{\delta}_1$, will be an unbiased estimator for the effect of the scholarship program on student outcomes. This coefficient of immediate relevance for literatures on scholarship and voucher programs.

The study is well-powered to detect even a small effect of the scholarship program. Assuming that the residual variance³⁷ in test scores is 0.7, the minimum detectable effect (MDE) for the PP_{govt} sample is 0.067 SDs ($\alpha = 0.05$ and $\beta = 0.8$). For the P_{govt} sample, the MDE is 0.064 SDs.

Treatment assignment – winning or losing a scholarship – did not completely determine a student’s choice of school. In particular, not all scholarship winners ultimately enrolled in a Bridge school, while some lottery non-winners did. For this reason, $\hat{\delta}_1$ may not capture the true effect of attending Bridge relative to other schooling options.

An estimate of the effect of attending Bridge is obtained by scaling the ITT estimate by an estimate of the share of students induced to switch into Bridge due to the scholarship lottery. This re-scaled effect can be interpreted as the local average treatment effect (LATE) on those who are induced to attend Bridge under the assumption that the scholarship itself does not influence students’ observed outcomes except through its effect on school choice.³⁸

Let $D_i \in \{0, 1\}$ indicate whether individual i attended a Bridge school at the end of 2017³⁹. Then the effect of winning the scholarship, Z_i , on Bridge attendance can be estimated by:

$$D_i = \pi_0 + \pi_1 Z_i + \eta_i . \tag{3}$$

The coefficient π_1 gives an estimate of the percentage of students who are induced to switch from other educational options into Bridge schools by the lottery. The effect of Bridge on

³⁷Controls for baseline characteristics are described in Section 9.5.

³⁸This assumption is common in the literature on school vouchers (Rouse (1998); Abdulkadiroğlu et al. (2018); Dean and Jayachandran (2019)). Section 9.3 describes empirical strategies to evaluate the plausibility of this assumption.

³⁹We will also explore estimates that replace this endogenous variable with a county of the number of years at Bridge.

students who are induced to switch into Bridge is then given by the Wald estimator:

$$\hat{\rho} = \frac{\hat{\delta}_1}{\hat{\pi}_1} . \tag{4}$$

An equivalent estimate can be obtained using 2SLS. The scholarship effect will be estimated separately for all the samples included in Table 12, which shows the estimated share of students who were induced to switch into Bridge schools by the scholarship.

Power simulations were conducted by generating a simulated outcome under the assumption that the residual variance of the outcome, after controlling for strata and other baseline controls, will be 0.7. The simulated outcome is then used to estimate a 2SLS specification of the effect of the number of years of attendance at a Bridge school. The results are presented in Table 21. The MDE represents the effect of attending Bridge in 2017. Each simulation is conducted for the sample of pupils who were followed at endline. Because this variable is highly collinear with a variable indicating that a pupil was enrolled in Bridge for two years, these estimates can be interpreted as representing the effect of attending Bridge for two years. Under an assumption of constant, non-decaying annual effects, an annual effect of attending Bridge can be obtained by dividing the resulting estimates by two.

Table 21 reports results for seven samples. As documented above, there are differences in counterfactuals and attrition results across these samples. Section 7 shows that these samples vary in terms of the usual counterfactual school environment for scholarship non-recipients. Section 6 shows that there is variation across samples in terms of differential attrition.⁴⁰ Larger samples without any evidence of differential attrition (for example PP_{gout} , P_{gout} , or their combination) may provide the greatest power to detect an effect, but readers with an interest in a particular choice margin or policy question may prefer to focus on other sub-samples for that reason.

⁴⁰Specifically, Sample PP_{priv} scholarship recipients were over 4 percent more likely to be followed at endline than non-recipients.

Table 21: Power simulation results

Sample	Reduced form				2SLS			
	SE	N	MDE		SE	N	MDE	
			$\beta = 0.8$	$\beta = 0.9$			$\beta = 0.8$	$\beta = 0.9$
$PP_{govt} + P_{govt}$	0.019	8,373	0.053	0.061	0.055	8,330	0.153	0.177
PP_{govt}	0.028	3,789	0.079	0.091	0.083	3,781	0.232	0.269
P_{govt}	0.025	4,584	0.071	0.082	0.072	4,549	0.203	0.234
$PP_{brig} + P_{brig}$	0.043	1,592	0.121	0.140	0.190	1,583	0.531	0.614
PP_{brig}	0.090	388	0.251	0.291	0.498	382	1.394	1.613
P_{brig}	0.050	1,204	0.139	0.161	0.203	1,201	0.568	0.658
P_{priv}	0.059	989	0.164	0.190	0.212	958	0.594	0.687

Notes: Each row represents the results of a power simulation using the sample of pupils followed at endline. Each simulation assumes that the residual variance of the outcome after controlling for strata and other covariates will be 0.7. Each simulation records the standard error resulting from running either a reduced form or 2SLS specification using the simulated outcome variable. In the case of 2SLS, the endogenous variable is an indicator for whether the pupil attended a Bridge school.

9.2 Decomposing the Scholarship LATE

The estimated effect of the scholarship program itself described in Section 9.1 compares the effect of a Bridge education relative to a composite of several existing alternative educational environments: government school, other private schools, and non-enrollment. It can also be interpreted as a weighted average of the effect of Bridge relative to each of these counterfactual educational environments.⁴¹ This section first describes the extent to which the scholarship offer affected choices across these schooling options – and, importantly, shows how they vary across schooling levels, and then presents several strategies for decomposing the scholarship effect.

The next subsection describes a strategy to decompose the LATE described in Section 9.1 into multiple component sub-LATEs representing the relative effect of Bridge for different counterfactual educational environments.

While it is technically possible to estimate all sub-LATEs corresponding to all three

⁴¹See Appendix Section D for a more detailed discussion of the derivation of this weighted average.

counterfactual education choices (government, non-Bridge private, and non-enrollment), this can lead to a loss in overall precision. In order to maximize the precision of the decomposed estimates, this study will focus on identifying a smaller number of sub-LATEs. As has been shown above, some sub-LATEs can be ignored because they play a quantitatively negligible role in the choices of students. In 2017, non-enrollment is rare and the effect of the scholarship on non-enrollment this sub-LATE is unlikely to have more than a negligible influence on the overall effect. Therefore, the sub-LATE for non-enrollment will not be included in the vector of endogenous variables. Including an endogenous variable with as little variation as non-enrollment in 2017 cannot produce results with a reasonable degree of precision.

Some readers may feel that the effect of the scholarship on non-enrollment should be considered in the decomposition of the main effect. To address those concerns, we will explore an additional specification that replaces the endogenous variables (attendance in 2017) with the pupil's attendance in 2016. As will be seen below in the results of power simulations, this specification is still underpowered to detect even large effects of non-enrollment; however, it will at least be powered to detect the kinds of extreme effects that would be necessary for non-enrollment to have a significant influence on the aggregate effect.

Based on qualitative observations by the research team and the results of the local education environment survey, it also not obvious that a clear distinction can be made between government and private provision of pre-primary education. Both government and private pre-primary programs require fee payments and there is no enforced common policy regarding teacher qualifications or how the schools should be administered. One reason that so many pre-primary students attend private schools in this setting is that government pre-primary has only recently been introduced. Historically, many private preschool providers were located in public schools, but did not receive public funding. While among primary schools, there seems to be a perception that private schools are of substantially higher quality and are less crowded than public schools, at the pre-primary level, there seems to be little basis to suspect that "government" and "private" provision are substantially different in terms of

value-added at the pre-primary level. For some respondents, the distinction between government and private programs may also be unclear at the pre-primary level. Programs may be supported by the community and located on the premises of public primary schools. These features of the environment motivate a comparison of Bridge to a combination of programs reported in the data as “government” and “private”.

In summary, the decomposition strategy that is described in the next subsection will distinguish between the following counterfactual educational environments:

- Bridge primary school;
- government primary school; and
- other, non-Bridge, private primary school.

Additionally, using attendance in 2016 as the basis for the endogenous variables, analogous tests will, for the sample of pre-primary students, distinguish between the following counterfactual education environments:

- Bridge pre-primary school;
- Other pre-primary school (either government, or other private); and
- Non-enrollment.

9.2.1 Decomposing the Scholarship LATE

As documented above, winning the scholarship lottery induced students to switch out of several education environments into Bridge. This section describes a strategy to further decompose the scholarship effect into sub-component effects capturing the relative impact of each environment (e.g. the effect of attending Bridge relative to a government primary school).

A common strategy for identification in this situation is to form multiple instruments using interactions of baseline covariates with treatment to use heterogeneity in first stage effects across individuals to identify the effects of multiple endogenous variables. Kling et al.

(2007), Kline and Walters (2016), and Kirkeboen et al. (2016) are a few recent studies to employ this strategy. Reardon and Raudenbush (2013) and Hull (2018) describe assumptions under which the overall effect can be decomposed. Hull describes the assumption concisely as one of homogeneity of LATE across the covariates used to form the interactions with treatment. Appendix D describes the problem formally, extending the set-up in Hull (2018) to a case where there are four counterfactuals/mediators.⁴²

In selecting baseline covariates to be used for this exercise, the research team has considered several factors. Instruments may vary in terms of the precision of the results and the point estimates obtained. Useful instruments must predict variation across individuals in the response to treatment. One strategy might be to use all covariates; however, using a very large set of covariate interactions can lead to biases due to many weak instruments (Bound et al. (1995)).

Five sets of instruments have been considered:

Set 1: A first set of instruments will consist of 10 covariates indicating the students baseline reported plans in the absence of the scholarship, current school type attended, socio-economic status, and how the applicant was recruited:⁴³.

- Enrolled in school at time of application (2015)
- Enrolled in NGO/private school at time of application
- Formerly a Bridge student
- Planning to attend a Bridge school at time of application
- Planning to attend a government school at the time of application
- Planning to attend an NGO or private school at the time of application.
- Primary caregiver’s occupation is agriculture
- Primary caregiver’s occupation is casual laborer
- Household monthly income (Kes) (inverse hyperbolic sign transformed)

⁴²This allows for the possibility that the scholarship has a direct effect on outcomes (i.e., the scholarship influences “always takers” who would have attended Bridge in the absence of the scholarship). This possibility will be addressed later in this document.

⁴³See Section 3.2 for details on recruitment procedures.

- Recruited by IPA

Set 2: This set will consist of four scalar predictions of the likelihood that a pupil would have, in the absence of the scholarship, enrolled in each of the four education options (government, Bridge, non-Bridge private, or unenrolled). There appears to be considerable variation in counterfactuals at the academy level. Fully interacting academy dummies with treatment would, however, risk bias due to over-fitting with many weak instruments. In order to make use of the variation at the academy level in counterfactuals for identification, we begin by estimating at the academy level, the probability that scholarship non-recipients are found in each of the education options. Following Abadie et al. (2018) this estimate is formed leaving individual i out of the academy level estimate. Finally, empirical Bayes shrinkage is applied to the leave-one-out estimator to minimize predicted mean square error. See Appendix F.2 for further details.

Set 3: This set will consist of interactions between treatment and 44 county dummy variables. This coarser level of geography may capture much of the local variation in counterfactuals with less risk of overfitting due to the use of many weak instruments.

Set 4 combines Set 1 and Set 2.

Set 5 combines Sets 1, 2, and 3.

Sets 3 through 5 contain many instruments, and the robustness of results using these instruments will be examined using JIVE and LIML.

9.2.2 Statistical Power for Decomposition

This section describes the results of power simulations using the data on choice of education environment. Minimum detectable effects were obtained by simulating an outcome with a standard deviation of 0.7, then running the 2SLS test and recording the standard errors.⁴⁴ The MDE is then 2.8 times the standard error of the resulting estimates.⁴⁵

⁴⁴The choice of 0.7 for the standard deviation of the simulated value was motivated by estimates of the residual variance in test scores after controlling for student and site covariates including baseline test scores, gender, and county fixed effects.

⁴⁵That is, $t_{1-\beta} + t_{\alpha/2} \approx 2.8$, for $\alpha = 0.05$ and $\beta = 0.8$.

In all cases, the endogenous variable is the number of years of exposure to each education environment.

Table 22 show MDEs and the Olev-Pflueger F-statistic from a specification with only a single endogenous variable. Andrews, Stock, and Sun (2018) recommend the use of this statistic for assessing the strength of instruments. Each F-statistic is from regression of the endogenous variable on the set of instruments. The existing literature has examined the performance of this F-statistic in detecting weak instruments problems in the case where the 2SLS specification includes a single endogenous variable. In this case, Andrews et al. (2019) show that F-statistics below 10 are problematic.

For pre-primary students, the MDE for the effect of being unenrolled relative to a Bridge school is greater than 1 SD for all sets of instruments and above 2 SDs in many cases. While a one SD effect from non-enrollment is plausible, this estimate is still very noisy and must be interpreted with caution. For primary school students in the PP_{govt} sample (Panel C), the MDE for the sub-LATE comparing Bridge attendance to other private school attendance is greater than 1 SD in all cases. Precision can be improved by including those students who were determined to be likely to enroll in other private schools in the absence of the scholarship (i.e. the P_{priv} subgroup). When this sample is included, the MDE for all instruments sets is below 0.9.

9.2.3 Plug-in Estimators

The framework described in the section above also yields a simple expression relating the “scholarship effect” (the LATE observed by instrumenting Bridge attendance with the scholarship offer) to the separate effects of the scholarship on each of the educational environments. The estimate $\hat{\rho}$ (from Equation 4), representing the effect of attending Bridge relative to a mixed counterfactual, is composed of a weighted average of the effects of attending Bridge relative to each of the relevant counterfactuals where the weights are composed of the effects of the scholarship offer on attendance at each of the counterfactual environments.

Table 22: Minimum detectable effects ($\alpha = 0.05, \beta = 0.8$) for decomposed sub-LATEs

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Sample PP_{govt}</i>					
Govt & Private	0.37 [44.4]	0.77 [82.2]	0.28 [13.8]	0.38 [32.2]	0.26 [10.4]
Unenrolled	2.05 [8.2]	5.23 [8.4]	1.42 [2.2]	2.09 [5.9]	1.10 [2.7]
<i>Panel B: Sample PP_{govt}</i>					
Government	0.93 [12.0]	1.29 [24.3]	0.55 [4.2]	0.81 [8.8]	0.51 [3.4]
Private	0.76 [18.8]	0.95 [45.7]	0.47 [6.4]	0.62 [15.3]	0.43 [5.1]
<i>Panel C: Primary P_{govt}</i>					
Government	0.38 [62.4]	0.30 [116.2]	0.23 [14.5]	0.27 [44.0]	0.23 [11.8]
Private	2.78 [3.6]	1.84 [8.4]	1.31 [1.6]	1.43 [4.6]	1.10 [1.7]
<i>Panel D: Samples P_{govt} & P_{priv}</i>					
Government	0.24 [54.8]	0.25 [130.8]	0.23 [13.7]	0.24 [40.8]	0.22 [11.1]
Private	0.87 [9.0]	0.90 [18.6]	0.86 [2.7]	0.81 [7.7]	0.68 [2.8]
<i>Panel E: Samples P_{govt} & P_{brig} & P_{priv}</i>					
Government	0.24 [58.0]	0.25 [155.0]	0.24 [14.2]	0.24 [44.0]	0.22 [12.4]
Private	0.86 [9.6]	0.88 [24.0]	0.87 [3.1]	0.80 [8.5]	0.68 [3.1]
<i>Instruments</i>					
10 covariates	X			X	X
4 LOO academy counterfactuals		X		X	X
44 County dummies			X		X

Notes: Within each panel, a row represents the minimum detectable effect for a sub-LATE comparing Bridge with the counterfactual educational environment indicated. Minimum detectable effects are obtained by simulating an outcome and then running the multiple-variable 2SLS specification. The MDE is the standard error of the resulting estimate multiplied by 2.8. Olea-Pflueger F-statistics in brackets. Panel A shows results using attendance in 2016. All other panels show results using attendance in 2017.

$$\hat{\rho} = - \sum_{j \in \{G, P, U\}} \frac{\hat{\pi}_j}{\hat{\pi}_1} \rho_j ,$$

where $\hat{\pi}_1$ is the effect of the scholarship offer on attendance at Bridge derived from estimating Equation 3, $\hat{\pi}_j$ indicates the first stage effect of the scholarship offer on attendance at each of the educational environments $j \in \{G, P, U\}$ representing government, non-Bridge private, and non-enrollment, and ρ_j represents the sub-LATE for those who are induced to attend Bridge who would have otherwise been in educational environment j . This expression is derived in Appendix D.

This expression shows that, with prior beliefs about ρ_k and ρ_l , one can estimate ρ_j for any $j \neq k \neq l$. For example, if someone believes, on the basis of previous studies, that most private schools are about 0.3 SDs better than government schools, and children who are unenrolled lose 1 SDs relative to where they would have been if they were enrolled, then $\rho_P = +0.3$ and $\rho_U = -1.0$, so rearranging the expression for the decomposition of $\hat{\rho}$ yields a point estimate of the sub-LATE comparing attendance at Bridge with attendance at a government school. Therefore, if all but one of the sub-LATEs is pinned down by assumption, the remaining sub-LATE is identified and can be estimated using the data in this study.

The research team will use results from the published literature on education, especially in Kenya, to define a range of possible values for sub-LATEs and then use these estimates to form bounds on the remaining sub-LATEs. Additionally, the research team will also compute the sub-LATE for ρ_j under the assumption that OLS estimates of the relationship between attendance at each educational environment and outcomes approximate the relevant sub-LATEs. These OLS estimates can be further refined using controlled specifications and Oster (2019) bounds. Furthermore, the results of the multiple endogenous variables IV estimation described in Section 9.2 will also provide guidance on reasonable values of each sub-LATE.

The published literature provides some guidance on the effects of private schools in Kenya

relative to government schools at the primary school level. Using non-experimental methods, Bold et al. (2013) estimate a 1 SD effect of private schools on KCPE test scores. If pupils graduating from private schools spent 8 years in those schools, then, under an assumption of constant linear effects, the annual average effect of private schools is approximately 0.125 SDs. We consider this effect in the context of a larger literature on vouchers and private schools that reveals mixed results. Some studies, such as Angrist et al. (2002) find an effect 0.2 SDs on test scores of a randomized multi-year voucher program in Colombia. But several other studies have failed to reject the null hypothesis of any effect. Hsieh and Urquiola (2006) find no effect of a voucher program in Chile using a similar identification strategy to Bold et al. (2013). Recent evidence from Muralidharan and Sundararaman (2015) find no effects from a recent voucher program in the Indian state of Andhra Pradesh, except on Hindi scores. They note that public schools in that context do not teach Hindi. It is also possible in the Kenyan context that the private school sector may have different effects on different subjects, although, it should be noted that, in the case of this study, all subject test scores are based on the official Kenyan curriculum. It should also be noted that some studies have found negative effects of voucher programs. Abdulkadiroğlu et al. (2018) found that a newly instituted voucher program in Louisiana reduced student achievement by 0.4 SDs on math scores and effects around -0.2 SDs on other subjects. Importantly, the authors speculate that the results could reflect adverse selection of private schools into the voucher program, so the observed effects may not be representative of private schools overall in that setting. Also, the program was being evaluated in its first year and part of the negative effect could be attributed to disruption effects from students switching schools. Finally, there is evidence that some public charter schools are effective at raising test scores (Abdulkadiroğlu et al. (2016)), so it is possible that the negative effect reflects the fact that many students would have been in these schools in the absence of the voucher. Still, while the Louisiana study may be exceptional, it serves as a reminder that there may be heterogeneity in the quality of instruction at private schools. If instructional quality is difficult to observe, or if

parents care about features of schools aside from instructional quality, private schools could have negative effects on student test scores. Given this evidence, the research team will consider effects in the range of -0.1 SDs to 0.2 SDs.

The published literature also provides some guidance on plausible effects of non-enrollment. It is worth noting that, in the present study, much of the effect of the scholarship on any school participation is concentrated among the pre-primary sample. The 4.7 percentage point effect of the scholarship on participation explains most of the 0.05 fewer years of school participation that scholarship pupils received. With this in mind, we first focus on understanding the plausible effects of a one year delay in school participation. Dean and Jayachandran (2019) evaluate a voucher program that induced higher kindergarten participation rates and find that, for those who are predicted to receive home care or enroll in Anganwadi centers (a form of day care) score higher on cognitive tests. They estimate using a decomposition method similar to that described in Section 9.2, that Hippocampus Learning Centers (HLC), a private provider of early childhood education, improve scores for students whose alternative is home care or Anganwadi by 1.4 SDs. However, they also find that students who are induced to switch out of other kindergarten programs into HLC score 0.4 SDs higher in their first year. Although they do not report estimates in this form, the difference between home care and Anganwadi on the one hand and non-HLC kindergarten options is approximately 1 SD. By the second year, however, this difference reverses sign, falling to approximately -0.1 SD.

It is common in the literature on early childhood education in rich countries to find moderate or large effects on test score outcomes of early childhood education that fade out rapidly. Results from the Head Start Impact Study (Puma et al. (2012); Walters (2015); Kline and Walters (2016)) find early effects of Head Start on test scores around 0.3 SDs that fade out rapidly. A recent large scale RCT evaluating Tennessee's state-wide pre-kindergarten program found mostly positive effects between 0.3 and 0.4 test scores, but found that these effects faded out almost completely by the end of kindergarten.

There is also some evidence on the effects of any school participation on test scores older children provides, some of it suggestive of large positive effects. Burde and Linden (2013) show that a program that provided local schools in Afghanistan had a large impact on school participation and large positive effects on test scores of 0.4 SDs for boys and 0.6 SDs for girls. Benhassine et al. (2015) evaluate an experiment that provided cash transfers to encourage school participation in Morocco. The intervention was found to have significant impacts on school participation for primary school students, but small and statistically insignificant effects on test score outcomes. It should be noted, however, that the lack of statistical significance may be a result of such interventions having only moderate first stage effects on school participation. Although the result is not reported, scaling the estimated effect on test scores by the “first stage” effect of the transfer on school participation would indicate a large effect of approximately 1 SD, although this effect would not be statistically different from zero.

With these results in mind, we will consider annual effects of being unenrolled in primary school between -0.5 SDs and 0.

9.3 Income effect of the scholarship

In the discussion above, it has been assumed that the scholarship did not influence outcomes except insofar as it had an effect on attendance at a Bridge school. This section describes strategies to test the assumption that the scholarship only affects pupils through its effect on school attendance. This section also discusses potential modifications to the plan of analysis in the event that tests indicate a violation of this assumption.

As described in Section 7.2 and Table 13, many pupils attended Bridge, even when they did not win the scholarship. In the pooled PP_{govt} and P_{govt} samples, non-recipients spent on average 0.472 academic years in Bridge schools.⁴⁶ In the PP_{brig} and P_{brig} samples pooled sample, non-recipients spent 1.438 years in Bridge schools. And in the P_{priv} sample,

⁴⁶Recall that the experiment covers the 2016 and 2017 academic (equivalently calendar) years, so this variable has a minimum of zero and a maximum of 2.

non-recipients spent 0.700 years in Bridge schools. For those students who would have attended Bridge in the absence of the scholarship, winning the scholarship represents an infra-marginal transfer to the household. This transfer may have an effect on the pupil. From the perspective of instrumental variables, such an effect would violate the exclusion restriction that the scholarship only influence the pupil's outcomes through its effect on enrollment.

To begin, the estimated effects of the scholarship on Bridge attendance for the PP_{govt} and P_{govt} samples (pupils who were relatively more likely to attend government schools in the absence of scholarship) will be compared to the corresponding effects of attending Bridge in the PP_{brig} and P_{brig} samples (pupils who were enrolled in Bridge at the time of the scholarship program). The income effect is most pronounced for those pupils in the PP_{brig} and P_{brig} samples, so, if the income effect is driving the effect observed in the PP_{govt} and P_{govt} samples, even larger effects should be observed in the PP_{brig} and P_{brig} samples.

To formalize this comparison, it is possible to redefine the endogenous variable so that, even if there are direct income effects of the scholarship on pupil outcomes, the exclusion restriction that the scholarship only affect the outcome through the endogenous variable is satisfied. Specifically, if the goal of 2SLS estimation is to obtain an estimate of the effect of attending Bridge *for free*, then the possibility of direct income effects of the scholarship is no problem for estimation.

Using the distinction between attendance at Bridge *for free* and attendance at Bridge *at cost*, it is possible to use the same approach to decompose the effect of the scholarship program into sub-LATEs representing attendance at different school types (government, private, unenrolled) as well as a sub-LATE representing attending Bridge *at cost*. Specifically, assume a vector of covariate controls C_i and a vector of C_i fully interacted with treatment $M_i(Z_i)$ are available.

It is possible to obtain separate estimates of the effect of each of the 4 counterfactuals using the following first stage specifications

Variable	Interpretation
$j \in \{G, B, P, U\}$	School Environment choice (endogenous variable)
G_i	Attendance at government school
B_i	Attendance at Bridge <i>at cost</i>
P_i	Attendance at non-Bridge Private school
U_i	Non-enrollment
s	Academy index

$$G_i = C_i\Gamma_1 + M(Z_i)\Gamma_2 + \epsilon_i$$

$$B_i = C_i\Omega_1 + M(Z_i)\Omega_2 + \eta_i$$

$$P_i = C_i\Pi_1 + M(Z_i)\Pi_2 + e_i$$

$$U_i = C_i\Upsilon_1 + M(Z_i)\Upsilon_2 + u_i$$

The second stage specification is then

$$y_i = \alpha + \hat{G}_i\gamma + \hat{B}_i\beta + \hat{P}_i\pi + \hat{U}_i\nu + C_iX_ia_i$$

Under the assumption that each sub-LATE, in expectation, is the same for pupils with different values of C_i , β captures the relative effectiveness of attending Bridge *at cost* compared to attending Bridge *for free*. The other coefficients are interpreted similarly. For example, γ indicates the relative effectiveness of government school and Bridge *for free*.

Table 23 shows the results of power simulations for a test of the relative effectiveness of Bridge *at cost* and all non-Bridge options (Government, non-Bridge private, and non-enrollment). As in Table 22, the power simulations are used to produce MDEs and Olea-Pflueger F-statistics. The effect on Bridge *at cost* represents the relative effectiveness of Bridge *at cost* and Bridge *for free*. It is also possible to further decompose the non-Bridge options. The MDEs for non-Bridge options when also estimating the effect of attending Bridge *at cost* are very similar (but almost uniformly marginally greater) than those reported in Table 22, so these results have been omitted in the interest of brevity.

Table 23: Minimum detectable effects ($\alpha = 0.05, \beta = 0.8$) for decomposed sub-LATEs

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: PP_{govt} sample</i>					
Anything but Bridge	0.60 [52.2]	2.77 [114.8]	0.50 [14.0]	0.66 [37.6]	0.35 [11.2]
Bridge at cost	0.70 [77.5]	3.29 [176.7]	0.61 [18.3]	0.77 [60.0]	0.40 [16.9]
<i>Panel B: $PP_{govt} + PP_{brig}$ sample</i>					
Anything but Bridge	0.43 [48.5]	0.55 [122.0]	0.49 [13.9]	0.44 [36.1]	0.36 [10.9]
Bridge at cost	0.41 [103.3]	0.51 [276.4]	0.50 [24.5]	0.40 [85.1]	0.34 [22.2]
<i>Panel C: P_{govt} sample</i>					
Anything but Bridge	0.77 [86.8]	0.65 [158.2]	0.30 [18.8]	0.58 [59.3]	0.28 [15.3]
Bridge at cost	1.49 [58.9]	1.22 [117.8]	0.53 [13.0]	1.08 [46.6]	0.49 [12.1]
<i>Panel D: $P_{govt} + P_{priv}$ sample</i>					
Anything but Bridge	0.42 [73.1]	0.41 [165.8]	0.30 [19.5]	0.39 [53.5]	0.29 [14.4]
Bridge at cost	0.66 [71.1]	0.61 [192.1]	0.47 [17.9]	0.58 [62.0]	0.43 [15.6]
<i>Panel E: $P_{govt} + P_{brig} + P_{priv}$ sample</i>					
Anything but Bridge	0.27 [75.8]	0.28 [196.7]	0.32 [20.2]	0.27 [57.5]	0.24 [16.2]
Bridge at cost	0.27 [171.1]	0.27 [525.1]	0.36 [34.8]	0.26 [151.0]	0.24 [39.0]
<i>Instruments</i>					
10 covariates	X			X	X
4 LOO academy counterfactuals		X		X	X
44 County dummies			X		X

Notes: Within each panel, a row represents the minimum detectable effect for a sub-LATE comparing Bridge with the counterfactual educational environment indicated. Minimum detectable effects are obtained by simulating an outcome and then running the multiple-variable 2SLS specification. The MDE is the standard error of the resulting estimate multiplied by 2.8. Oleva-Pflueger F-statistics in brackets.

The results of this analysis will be considered alongside other results related to the question of whether the scholarship may have had an effect on pupils beyond its effect on attendance. These considerations are discussed in Section 11.3. If there is evidence that the scholarship had an effect on pupils beyond its effect on attendance, the interpretation of all tests will need to be modified to account for this fact.

9.3.1 Plug-in Estimators for direct effect of the scholarship

As discussed in Section 9.2.3, it is possible to construct bounds on a given sub-LATE using information on the effect of the scholarship on education options, the ITT effect of the program, and priors about the range of possible values for some of the sub-LATEs. We will extend this approach to the problem when we allow for a direct effect of the scholarship on test scores.

The literature on cash transfers for low income households provides some guidance on expected effects of income on student test scores. Benhassine et al. (2015), noted above, find no effects of either conditional or labeled cash transfers on student test scores. Therefore, if we attribute their entire observed (and statistically insignificant) effect on test scores to the direct effect of the cash transfer, implicitly assuming that school participation has a negligible effect on student test scores, then the direct effect of the cash transfer on student test scores was approximately 0.07 SDs, and not statistically different from zero.⁴⁷

The literature on cash transfers also points to the possibility that while there may be important income effects of the voucher on the household, the benefit of the income is unlikely to be concentrated entirely on the child to whom the voucher is assigned. For example, Behrman and Parker (2013) provide evidence that the income effects of a cash transfer may be improve the health of elderly household members. Haushofer and Shapiro (2016), evaluating an experiment that provided unconditional cash transfers in rural Kenya,

⁴⁷Transfers amounts varied according the age of the child. The available transfer amounts varied from 600 to 1000 MAD. They report that the average education expenditure in the control group was 180 MAD. Therefore, the transfer resulted in an income transfer for many households.

find small effects of large cash transfers on an education index⁴⁸, but large effects on the purchase of assets, durable goods, and a food security index. Even if households could purchase goods or services that would improve one child’s test scores, it seems that an extra shilling is unlikely to be spent exclusively to advance a single child’s education.

We also consider the possibility that the scholarship itself could undermine learning. First, pupils who receive a scholarship could be more or less motivated than pupils who do not receive the scholarship. Students who have not received a scholarship may work harder in the hopes of receiving a future merit-based scholarship. Kremer et al. (2009) find large effects of offering a merit-based scholarship in Western Kenya between 0.1 and 0.3 SDs. Another reason why the scholarship itself could have a negative effect on students who were going to attend Bridge anyway is if teachers feel less accountable to the families of children who are on scholarship.

With these results in mind, the research team feels that the common practice in the voucher literature in assuming that the direct effect of the voucher is negligible is a reasonable starting place. We will also consider direct effects of the scholarship on student test scores between -0.1 and 0.1. However, unless other empirical results from this study indicate the presence of a direct effect of the scholarship⁴⁹ the research team feels that, as is common in the literature evaluating voucher programs, it is reasonable to assume that the direct effect of the scholarship on students is negligible.

9.4 Interpretation of effect and alternative specifications

The 2SLS specifications described above employ binary endogenous variables to identify an effect of attendance at Bridge (and other counterfactual environments) at different points in time. Presumably, if there is a positive or negative effect of one education environment, it is likely that that effect is the result of an accumulation of smaller effects over time. It may

⁴⁸The education index includes enrollment, missing class, and expenditures. It does not include test scores.

⁴⁹We will consider specifically the results of the decomposition tests in Table 23 and the results of tests of effects of Bridge on siblings who do not win the scholarship but who were induced to enroll in Bridge because a sibling won a scholarship (see Section 11.3.3).

be of interest in some policy problems to understand how these effects relate to effects for different dosages of exposure to different educational environments.

The relationship between attendance at Bridge, government and private schools in any year of the program (2016 and 2017) is highly predictive of attendance in the other year. The scholarship increases the likelihood that a student spends both years at Bridge by 38 percentage points. Table 12 shows that the effect of the scholarship on 2016 Bridge attendance is 37 percentage points, and the effect on 2017 attendance is 35 percentage points. This supports the view that the scholarship's effect on Bridge attendance in either 2016 or 2017 is highly collinear with 2 years of Bridge attendance. These results suggest that the effect of the scholarship on the number of years that a student attends a given type of school is approximately twice the effect of the scholarship on the number of years in that school type.

If one assumes that the effect of Bridge is constant and past effects do not decay rapidly over time, the effects obtained in this section can simply be divided by two to obtain approximate annual effects of each school type. While this is a strong assumption, over the short term, it may be a reasonable approximation.⁵⁰

The effect of the scholarship on attendance at different school types is not constant over time in some cases. First, the effect of the scholarship on attendance among students in the *brig* samples is initially about half the size of the effect in 2017. This means that for about half of the students induced to stay in Bridge for longer, were in Bridge in 2016 and so received one year of Bridge education during the program. Even under the assumptions above, the estimated 2SLS effect for this sample may not represent the effect of two years

⁵⁰It is worth noting that the early childhood education literature documents rapid fade-out from preschool interventions and this may be viewed as evidence against the hypothesis that effects do not decay rapidly. However, this literature examines the effects of school versus home care and there are several reasons why this may not be informative when comparing the effects of different school options. One explanation for rapid fade-out in those studies is likely to be the result of catch-up among the control group in pre-school interventions as nearly all of the children in these studies all eventually enter school (Gibbs et al. (2013) speculate that this could be one source of fade-out for Head Start). There may be large effects in the first year of education. When comparing students who attended two separate learning environments for two years, there may also be dynamic complementarities between the quality of the learning environment that partially offset any decay in prior year effects.

of exposure. Second, for the PP_{govt} sample, the scholarship induced a small but significant share of students (approximately 4.7 percent) to enroll in school who likely would have been unenrolled.

One way to address these issues is to replace the endogenous attendance variables with the number of years attendance at each school type. Specifications using the number of years of attendance at each school type yield the average causal response which, under the same assumptions of constant, non-decaying effects can be interpreted as annual effects. We will report results from specifications analogous to those described in this section using an integer count of the number of years of each attendance type.

9.5 Controls

This section describes the additional baseline controls that will be included in the student level results.

Baseline specifications will include the following:

- Randomization strata dummies.⁵¹
- Grade in which the applicant planned to enroll in 2016 at the time of application dummies,
- A vector of baseline test scores
 - Self-reported term 3 end-of-term English and Math scores
 - Score on Uwezo test [students enrolled through IPA only]
 - Score on self-regulation test [students enrolled through IPA only]
 - Score on Ravens matrices assessment [students enrolled through IPA only]
 - MDAT scores [students enrolled through IPA only]
- Residual height (controlling for age)
- Gender
- Caregiver literacy
- Caregiver occupation dummies.

⁵¹In cases where leave-one-out predictions are used, strata dummies will be replaced with a treatment propensity score representing the mean of treatment within the strata cell.

When there are missing values for any of the controls, the missing value will be imputed to be the mean of the variable for non-missing observations and an indicator for missingness will be included in the vector of controls. This imputation procedure for control variables has been used elsewhere.⁵²

10 Outcomes

This section describes the outcomes that will be examined and reported, and hypotheses that will be tested. The analysis will examine how attendance at Bridge affects test scores and other outcomes. This section begins by describing the formation of academic subject test score outcomes that will be used to assess student learning about subject-specific material. After the discussion of academic subject test scores, this section will describe other data that will be included in the analysis.

10.1 Academic subject test scores

Subject test scores are commonly used as outcomes in the analysis of school value-added and they are an important outcome for many parents who are considering where to enroll their child and for many policymakers. Academic subject knowledge is closely related to preparation for the high-stakes primary school leaving exam, the KCPE. As described in Section 4, at endline students in ECD were tested in reading (both Kiswahili and English) and math; while students at the primary level were tested in Math, English, Kiswahili, Social Studies, and Science.

The academic subject knowledge tests were designed by consultants with teaching and curriculum development skills and knowledge in Kenya, and were based on the Kenyan national curriculum. See Appendix C for additional details.

Students were administered exams corresponding to the grade they were in at the end

⁵²See Abdulkadiroğlu et al. (2019)

of the 2017 academic year. Subject tests varied across grades, but tests for adjacent grades contained overlapping questions. These items allow for comparison across tests.

10.1.1 Item Response Theory (IRT) Estimation and Test Equating

For each subject s and class c , a two parameter (2PL) IRT model is estimated yielding a measure of “ability”, θ_i^{cs} for each student who sits for the class c battery of assessments.⁵³ 2PL IRT methods also yield estimates for each item in each test: items are characterized by b_k^{cs} , an estimate of the “difficulty” or location of the item k , and a_k^{cs} , an estimate of the “discrimination” of the parameter.

Prior to estimating the IRT model, items are dropped whenever a correct response is not predictive of correct responses on other items within the test. Specifically, for each item k on subject test s for class c , a count of correct items within the test r_{-k}^{cs} is formed. r_{-k}^{cs} is then regressed on an indicator for a correct response on item k . Items with a z-score below 1 are excluded. This test identifies two items to be excluded. We describe the items below.

One item, included in both the standard 3 and standard 4 English assessment, was dropped (item labels s3_e_21, s4_e_16). This item asked pupils

[s3_e_21/s4_e_16] Which one is correct?

1. All my friends **want** to play a game.
2. All my friends **wants** to play a game.

This item is negatively correlated with the score on the remaining items with a t-statistic of -0.98 for standard 3 students and -1.28 for standard 4 students. 44 percent of students responded correctly in standard 3 and 38 percent of students responded correctly in standard 4.

Another item was dropped from the Standard 2 science and social studies assessment. This question asked pupils

⁵³For a concise description of IRT methods used here, see Kolen and Brennan (2014), Chapter 6.

[s2_s_18] Which of the statements is not true about a compass needle?

1. The needle always points to the South
2. The needle always points to the North
3. The needle is magnetic
4. The needle is metallic

Only 20 percent of standard 2 respondents answered this item correctly, and the item was weakly correlated with the score on the remaining items with a t-statistic of 0.78.⁵⁴

10.1.2 Vertical Scaling

IRT test equating procedures will be used to place scores on a common scale. Several procedures for vertical scaling of tests administered to populations with different levels of ability are described in Kolen and Brennan (2014) (Chapter 6). Kolen and Brennan (2014) review literature that suggests that, for dichotomous IRT models, characteristic curve methods, such as Stocking and Lord (1983), produce more stable results than moment-based methods such as mean-sigma and mean-mean. They also review literature that suggests that concurrent estimation may be more reliable than item characteristic procedures when the assumptions for IRT are satisfied, but that item characteristic approaches such as Stocking-Lord may be more robust to violations of IRT assumptions. Considering the literature as a whole, we feel that moment-based procedures are dominated by item characteristic curve approaches and concurrent estimation, but that both item characteristic approaches and concurrent estimation should be considered.

We adopt the Stocking-Lord equating method as our preferred method of vertically scaling the subject tests. The resulting equating coefficients are reported in Table G2.

⁵⁴This item was also asked on the Standard 3 and Standard 4 exams where the item was estimated to be one of the most difficult questions. 25 percent of standard 3 students answered the question correctly and 30 percent of standard 4 students answered the item correctly. This pattern of results seemed to confirm that the question was too difficult for Standard 2 students and that correct responses mostly indicate noise. The item was not, however, removed from the Standard 3 and 4 exam.

As a robustness check, we will also create alternate scaled outcomes measures using a novel approach. This approach uses concurrent estimation for each pair of adjacent assessment levels to estimate equating coefficients that are then used to form chain equating coefficients. Specifically, for each subject assessment s , adjacent grades ($c - 1$ and c) we perform the following procedure:

1. Estimate a 2PL IRT model for $c - 1$ and c , concurrently;
2. Standardize the resulting ability estimates to the class $c - 1$ distribution; and
3. Estimate $x_{c-1,c}$ as the standard deviation of the standardized ability estimate for grade c .
4. Estimate $y_{c-1,c}$ as the mean of the standardized ability estimate for grade c

The results from this concurrent estimation strategy are reported in Table G3. The results show that Stocking-Lord and concurrent estimation yield very similar results.

A choice must be made of which scale all scores should be expressed in. We have chosen to standardize all scores in terms of the Standard 4 scale.⁵⁵

10.1.3 Composite Subject Knowledge Indices

The approach described above yields individual subject scores. In the interest of constructing a single achievement score that will serve as the focal measure of subject knowledge, individual subject scores will be standardized and combined into a an index following the approach of Kling et al. (2007).

This score will combine English and Kiswahili scores into a single “Language” index, then standardize the “Language” index and combine this with the standardized Math score. Many studies of the effects of educational interventions focus on Language and Math skills, and therefore this measure will be more broadly comparable to other related studies. There is much less evidence on what interventions influence knowledge about Science and Social

⁵⁵Reported results will also include the equating parameters discussed above that will allow for translation of results into the scales of other grades. This may be important when comparing effect sizes to other interventions for specific age groups.

Studies. For this reason, the science and social studies test will not be included in the subject knowledge aggregate index.

Results for individual subject tests, including science and social studies, will also be reported.

10.1.4 Testing for Robustness to Equating Method

Errors in the estimation of equating coefficients are unlikely to influence the results given the relatively small effects of treatment on grade progression. This is especially true for learning aggregates that incorporate a larger number of individual equating coefficients.

However, out of caution, the research team will conduct several robustness checks in order to assess whether differential grade progression or the choice of test equating procedure could influence the results.

First, estimates of the effect of the scholarship will be obtained using multiple test equating procedures. As noted above, we will use both the Stocking-Lord equating procedure and the concurrent estimation procedure.

Second, estimates of the effect of the scholarship will be obtained on sub-samples where the scholarship had a smaller effect on grade progression. Differential grade progression does not affect pre-primary students in the PP_{govt} sample because they all take the same test. Students projected to be in higher grades in primary school have much less differential grade progression. We will therefore consider analyses that restrict to these samples. Additionally, we have found that students who report higher baseline (Term 2 end of term scores) are less likely to be held back, and, therefore, less likely to have their assessment level influenced by treatment. An additional test will restrict to students with above median (within their geographic stratum) baseline scores. Obviously, it is possible that the effect will be different across any subset of the data due to heterogeneity in the treatment effect, so none of these robustness checks can be seen individually as yielding dispositive evidence for or against the validity of the test equating procedures described above.

10.1.5 Draw a person test

The ECD language assessment also include a component on related to the “Draw a person test”. For this item pupils were asked to draw someone standing up. The drawing was assessed by the field officers according to seven individual criteria, including whether the person has a head, body, arms, legs, hands or feet, one recognizable feature of the face, and two recognizable features of the face. These six items will be scored using IRT to form a scalar measure of achievement on the test.

10.1.6 Local content

One concern raised by the prospect of developing instructional materials internationally is that international experts may lack expertise or may de-emphasize local content. Building local content knowledge may be an important objective of public education if it is a component of a sense of shared national identity and citizenship.

This study will attempt to partially address this question of whether the scholarship helped or harmed local content knowledge.

We will identify items in the science and social studies subject test that are unlikely to be found on subject tests for similar age groups outside of East Africa. These items include material about local geography, industry, civics, and health. Because there are no items covering these subjects in assessments for ECD to Standard 3, this analysis will include pupils in Standard 4 and above. Starting in Standard 4, there are between four and eight items covering these subjects on each assessment. The items identified to be the basis of this local content knowledge score are described in Appendix Table G4. The score will be computed using the same IRT test equating methods described above. Robustness of the result to removing certain types of the items will also be considered. Specifically, some items deal with knowledge of infectious disease which may be viewed as priorities of the international development community and may not be as relevant to the questions about a

shared national identity.⁵⁶

10.2 Pupil Level Outcomes

We will test the null hypothesis that the scholarship did not affect any of the following outcomes listed in Table 24.

⁵⁶We will also examine whether the scholarship increased knowledge as measured by items relating to infectious disease.

Outcome	Grades	Notes	Survey
Subject knowledge tests			
Subject Knowledge Index	All	See Section 10.1.3.	Pupil
Language index	All	See Section 10.1.3.	Pupil
Mathematics IRT score	All	See Section 10.1.1.	Pupil
English IRT score	All	See Section 10.1.1.	Pupil
Kiswahili IRT score	All	See Section 10.1.1.	Pupil
Science and social studies IRT score	P	See Section 10.1.1.	Pupil
Local content knowledge	≥STD 3	See Section 10.1.6.	Pupil
KCPE test scores (self-reported)	P	The 2017 phone call survey, the endline survey, and an ongoing tracking survey with pupils have all collected this information. We will also test for an effect of the scholarship on test taking and on receiving above a 250.	Multiple
KCPE test scores (official)	P	Data have also collected on the location at which the student took the test and the name under which the student was registered. These data can be used to match students to public records of KCPE test scores, although the team does not currently have access to the official records.	Multiple
Cognitive and non-cognitive tests			
Draw a Person test	PP	See Section 10.1.5.	Pupil
Raven's Matrices	All	This task asked pupils to respond to 15 Raven's matrix items. The matrices became increasing difficult, and if a respondent missed more than 2 in a row, the task ended.	Pupil
Digit Span Recall	All	This tasks asks pupils to repeat a string of digits. The task includes eight items. The length of the the string of digits were between four and seven .	Pupil

Outcome	Grades	Notes	Survey
Head and Knees Task	PP	This is a measure of behavioral self-regulation (ECD only). The task begins with a practice round in which the child is asked to touch their head and their knees. After that, the field officer asks the child to do the opposite of what they are asked. The field officer records that either the child responds correctly, incorrectly, or self-corrects.	Pupil
Receptive vocabulary Task	≤STD 3	This task measures a pupil's receptive vocabulary. Pupils are shown a page containing four drawings depicting distinct activities, objects, shapes, etc. For each item, a trained field officer asks the pupil to point to the image on the page depicting a particular activity, object, shape, etc. The field officer indicates if the pupil correctly identified the image corresponding to the activity, object, shape, etc. The task is composed of two sections, the first in English and the second in Kiswahili. The task consists of 24 items.	Pupil
Torrance Test (Spoon task)	All	The task is based on the Torrance test. Pupils are presented with a spoon and asked to enumerate uses of the spoon. A trained field officer classifies these responses as either "normal" or "unusual". The number of unusual responses is summed to create the pupil's score. Assessor fixed effects will be included in the specification	
Academic Achievement			
Grade level	All	These outcomes have already been studied and preliminary results are presented in Section 7.5. Additional results will be examined using data from phone call tracking surveys in 2019.	Multiple

Outcome	Grades	Notes	Survey
KCPE test taking	P	An binary variable will be formed using data on KCPE test results to show that a household registered and took the KCPE exam.	Multiple
On time KCPE test taking	P	For each pupil a projected KCPE test taking year will be identified using baseline grade and assuming normal progression. A binary variable will be formed indicating whether a took the KCPE by the projected date. The variable will use multiple data sources including the 2017 phone call tracking survey, the end-line survey, and the 2019 phone call tracking survey. If data is missing for all surveys after the projected KCPE test taking date, this variable will be missing. This test is intended to examine whether the scholarship reduced the practice of holding back pupils ahead of the KCPE test either to prepare academically for the test or to save for secondary school fees.	Multiple
Secondary school transition	P	A binary variable will be formed using data from the endline survey and phone call tracking survey to indicate whether the pupil successfully transitioned to secondary school.	Multiple
Occupational Aspirations			
Aspiration to work in particular occupation category	All	[professional/manual]	Pupil
Aspiration to be particular occupation	All	[Doctor/Teacher/Driver/professional/white collar]	Pupil
Preference for female occupation	All	Using the sample of control students, the share of female pupils who aspire to occupation x is estimated, excluding individual i .	Pupil

Outcome	Grades	Notes	Survey
Preference for high achievement occupation	All	Using the sample of control students, the average test score of pupils who aspire to occupation x is estimated, excluding individual i .	Pupil
Homework and studying			
Homework assigned	All	Response to question: "During a typical week during the last school term in 2017, how many days did you get homework?"	Pupil
Typical weekly study hours	\geq STD 4	Response to question: "On a typical school day last school term in 2017, how many hours did you spend studying or working on homework outside of class?"	Pupil
Pupil had a quiet place at home to do homework.	All	Pupil affirms the following statement: "You have a quiet place to do homework at home."	Pupil
Pupil reported attendance at tutoring sessions	All		Pupil
Self assessed homework load relative to other students in school	All	"Last year, you got assigned more homework than other students in your school."	Pupil
Self assessed homework load relative to other schools	All	"Last year, you got assigned more homework than other students in other schools."	Pupil
Pupil social attitudes			
Gender bias	All	If pupil agrees with the statement "Boys are smarter than girls", then this indicator takes a value of 1	Pupil
Intra-ethnic group trust	All	If pupil answers affirmatively to the statement "In general, can you trust members of your tribe?"	Pupil

Outcome	Grades	Notes	Survey
Inter-ethnic group trust	All	If pupil answers affirmatively to the statement “In general, can you trust people in other tribes?”. If the results indicate that intra-ethnic group trust is unaffected by the scholarship, we will also estimate specifications that treat intra-ethnic group trust as a control variable.	Pupil
Democratic support	All	If agrees with the statement “In Kenya, elections (or voting) is the best way to choose our leaders (such as the president, governor, or MCA)” this indicator is coded as 1	Pupil
Parental engagement		Parental engagement will be measured by an index formed by standardizing and taking the mean of the following survey items:	
School committee	All	Someone in the HH has been a member of the school committee	Caregiver
Knowledge of pupil class rank	All	Caregiver reports knowing pupil’s class rank	Caregiver
Knowledge of pupil test scores	All	Caregiver reports knowing pupil’s final test scores	Caregiver
Caregiver-teacher have met	All	Caregiver has met pupil’s teacher	Caregiver
Caregiver-teacher meetings	All	Number of meetings with pupil’s teacher	Caregiver
Caregiver-head teacher have met	All	Caregiver has met pupil’s head teacher	Caregiver
Caregiver-head teacher meetings	All	Number of meetings with pupil’s head teacher	Caregiver
Parental satisfaction and complaints		(Note: complaints may be an indication of greater engagement or lower satisfaction.)	
Satisfaction	All	Parent satisfaction with pupil’s school	Caregiver
Complaints	All	Parent has lodged a complaint at school	Caregiver
Complaints about fees	All	Parent has complained about [fees/class size/teacher quality/teacher absences]	Caregiver

Outcome	Grades	Notes	Survey
Missed class			
Ever absent	All	Did the pupil miss any class in term 3 of 2017	Pupil
Total absences	All	Number of days the pupil missed class in term 3 of 2017	Pupil
Ever absent for fees	All	Pupil missed class interm 3 of 2017 because of lack of fees, uniform, or other fee or in-kind related payment	Pupil
Missed more than one week for fees	All	Missed more than a week due to missed fees	Pupil
Missed more than two weeks for fees	All	Missed more than 2 weeks due to missed fees	Pupil
Education expenditures and other inputs			
Total education expenditure	All	Kenyan shillings	Caregiver
Tutoring costs	All	Item is not available if caregiver was only able to give total expenditure amount	Caregiver
Tuition and Fees	All	Item is not available if caregiver was only able to give total expenditure amount	Caregiver
Additional tutoring sessions	All	Did the pupil take any additional tutoring sessions in 2017?	Caregiver
Additional tutoring sessions	All	Number of tutoring sessions in a usual week	Caregiver
Testing conditions			
		(Note: If there are observed differences in testing conditions, these variables may be included as controls in robustness checks of results.)	
Pupil had a meal today?	All	Controls for survey hour of day will be included with any specification testing for effects on this outcome. This will be used to assess whether there were observable differences in testing conditions.	Pupil

Outcome	Grades	Notes	Survey
Hours since last meal	All	Controls for survey hour of day will be included with any specification testing for effects on this outcome. This will be used to assess whether there were observable differences in testing conditions.	Pupil
FO reported testing conditions	All	This will be used to assess whether there were any observable differences in testing conditions.	Pupil
Pupil classroom experience			
School has latrines	All	Pupil reported	Pupil
Latrine cleanliness	All	Latrines at school are clean	Pupil
Library	All	School has a library	Pupil
Computer access	All	School has computers or tablets for pupils to use	Pupil
Computer usage	All	Pupil has used a computer or tablet in school	Pupil
Playing area	All	Is there an outside playing area	Pupil
Safety of playing area	All	Are there any hazardous objects in the field that could hurt someone?	Pupil
Fence	All	Is there a wall or fence surrounding the school's grounds	Pupil
Start time	All	Time of day at which school starts	Pupil
End time	All	Time of day at which school ends	Pupil
Saturday school	All	Pupil has ever attended school on a Saturday	Pupil
Teacher adaptation (struggling)	All	The question asks: If pupil were falling behind, do teachers give extra help? This item is asked on both the parent and pupil survey. Results will be reported for both separately.	Multiple

Outcome	Grades	Notes	Survey
Teacher adaptation (excelling)	All	The question asks: If pupil were ahead, would teachers give extra help? The question is asked of caregivers and pupils. Results will be reported for each separately.	Multiple
Adaptation index	All	An index will be computed that averages standardized responses on the pupil and caregiver survey. The four items would include two from each of the pupil and caregiver survey: Teacher adaptation (struggling) and Teacher adaptation (excelling)	Multiple
Language of instruction	All	Teacher uses a different language than English or Kiswahili	Pupil
Overcrowding			
Availability of desks and chairs	All	Some children in pupil's class do not have a desk or chair	Pupil
Noise	All	Outside noises make it difficult to hear teacher	Pupil
Interference from other students	All	Other classmates bother pupil	Pupil
Fights	All	Sometimes classmates fight in pupil's class	Pupil
Textbooks			
At least one subject with textbook	All	Pupil had at least one subject with a textbook	Pupil
At least two subjects with textbook	All	Pupil had at least two subjects with a textbook	Pupil
Number of textbooks	All	The number of subjects for which a pupil has textbooks	Pupil
Subject textbooks	All	Individual binary variables indicating that a pupil has a textbook in Mathematics, English, Kiswahili, and Science and Social Studies	Pupil

Outcome	Grades	Notes	Survey
Questions and Help			
Treatment when pupil needs help	All	Pupil likes the way their teacher treats them when they need help	Pupil
Treatment when pupils ask questions	All	Teachers were nice to pupil when they asked questions	Pupil
Teachers ask questions to gauge understanding	All	Pupil's teachers asked questions to be sure they were following along when they were teaching	Pupil
Adaptation in instruction	All	If pupil did not understand something, your teachers explain it another way	Pupil
Teachers ask questions	All	Teachers [Sometimes/Often] ask questions to students in class	Pupil
Teachers only ask questions of certain students	All	Pupil feels that teachers only asked certain students questions	Pupil
Students ask questions in class	All	Students [sometimes/often] ask questions in class	Pupil
Helpfulness of grading	All	When pupil's teachers grade homework, they wrote on pupil's papers to help them understand	Pupil
Play time	All	Teachers let pupil play a lot at school	Pupil
Testing			
Time spent on exams	All	Pupil feels that they spent a lot of time taking exams	Pupil
Exam preparation 1	All	"Last year, a lot of class time was spent on getting you prepared to take a test"	Pupil
Exam preparation 2	All	"Last year, did your teachers help you prepare for a test like the one you took today?"	Pupil
Corporal punishment			

Outcome	Grades	Notes	Survey
Harsh language	All	Teachers use bad/harsh/threatening words or insult students (pupil reported)	Pupil
Teachers carry pipes, canes, or sticks	All	Teachers carry pipes, canes, or sticks	Pupil
Hitting	All	Teachers hit, pinch, or slap students	Pupil
Public recognition	All	Teachers use public recognition to reward good work	Pupil
Teacher absences			
Never absent	All	In a usual week, teacher is never absent from class	Pupil
Absent less than 4 hours per week	All	In a usual week, teacher is absent from class less than 4 hours total	Pupil
Absent less than 8 hours per week	All	In a usual week, teacher is absent from class less than 8 hours total	Pupil
Absent more than half the week	All	In a usual week, teacher is absent more than half the week or more	Pupil
Teachers leave class alone for long periods	All	In a usual week, teacher left the class alone for a long time [Sometimes/Often]	Pupil

11 Interpretation and Extensions of Results

This section describes the context of the study and how this context motivates hypotheses to be tested and their interpretation. First, the model employed by Bridge will be described with particular emphasis on why a Bridge education may differ from other options available to applicants to the scholarship program. Next, we will describe several important themes that will inform the interpretation of the results. These include whether the scholarship may have had a direct income effect on those pupils who would have enrolled in Bridge even if they had not received the scholarship. Then, we will discuss how the results can be informative about other mechanisms that might explain observed effects.

11.1 The Bridge model

The model used at Bridge schools to deliver education may differ considerably from models employed in other schools in this context. Reviewing these differences in educational/instructional approaches will guide the interpretation of some results, especially those concerning mechanisms driving any observed effects.

Bridge teachers use detailed lesson plans (sometimes referred to as scripts) in their instruction. Scripts are created centrally and used by teachers via tablet e-readers. Bridge states that these lessons are designed to follow the national curriculum, but some stakeholders have raised concerns that these materials are not well aligned with the curriculum. Lesson completion data is collected using the tablet, providing additional monitoring of teacher effort.⁵⁷

Teachers at Bridge schools may have less formal teacher training and fewer academic credentials than teachers in many other schools in their area. Section 5 shows that Bridge teachers tend to have less education and have been working at their schools for shorter

⁵⁷It is important to note, however, that while data may be collected, it is not necessarily of high quality or practical use for monitoring.

durations. Although teacher recruiting practices have changed in recent years, during the time period under study, Bridge recruited teachers who were not necessarily trained as teachers and many only had a secondary school degree.

The centralized development and dissemination of lesson plans standardizes a key component of educational production that is typically performed by individual teachers in other schools. One view is that some portion of the work of teachers can be completed centrally at high quality, and that teachers with relatively low levels of formal education or without a teacher training college degree can deliver these lessons effectively. In interpreting the results of the tests above, the research team will consider whether the results are consistent with the view that Bridge provides a product with uniform quality and that is robust to variation in teacher characteristics. The research team will consider several results in order to understand whether evidence is consistent with the story of standardization and technological substitution, including comparing the experimental results to descriptive results on the level of teacher skills, examining the effect of the scholarship program on the dispersion of student test scores, and testing for the presence of heterogeneity in the effectiveness of Bridge across sites.

In interpreting the results described above, it is important to keep in mind other differences between the Bridge model and other schools. Recruitment and training practices may be important parts of the Bridge model. The application process for teachers includes testing on subject knowledge (in math, reading, and Kiswahili), pedagogical knowledge, and technology skills. Once recruited, Bridge teachers receive additional training. According to data provided by Bridge, teachers in Bridge academies in Kenya receive approximately 150 hours of training. Once active, teachers are supposed to receive additional on-the-job training in the form of feedback from academy managers and central Bridge staff. Classroom observations are common in Bridge schools. Academy managers are tasked with carrying out regular classroom observations and providing teachers with feedback. Tables 3, 4, and 5 show that over 70 percent of academy managers at Bridge report regularly conducting

classroom observations. Some classroom observations are administered centrally, in addition to regular classroom observations conducted by academy managers. 4,577 central classroom observations were conducted in Bridge's Kenyan academies in 2017 according to data provided by Bridge, an average of more than one per classroom. The SDI survey shows that only around 40 percent of government school head teachers regularly conduct classroom observations. Government schools also regularly are inspected by quality assurance officers from the Ministry of Education. Over 80 percent of government schools in the local education environment survey reported a visited from a QA officer within a year of the survey (in 2012).

Teacher absence monitoring in Bridge schools uses a combination of on-the-ground reports from the academy manager and monitoring through the tablet. Bridge also uses substitute teachers, so recorded teacher absences may not reflect pauses in classroom progress. The data also record unexcused absences which account for about 11 percent of all absences. Recorded absences are classified in Bridge data and include absences due to maternity, compassionate, and sick leave.

In addition to teacher absence, private schools, including Bridge may differ from government schools in terms of the amount of class time that pupils miss due to an inability to pay fees. Although it is possible that pupils miss classes in both government and private schools, the higher fees in private schools may result in more missed class. Student absences may also interact with teacher absences. If teachers are more likely to be present and teaching in private schools on any given day, then a pupil who misses class in a private school may fall further behind his peers than another pupil missing the same amount of class in a government school.

Additionally, Bridge and other private schools may provide more instructional time, and this greater instructional time could substitute for other inputs including teacher qualifications. Tables 3, 4, and 5 show that Bridge schools days are longer than both nearby government and nearby private schools.

One reason given for the growth of private primary schools in Kenya is that, since primary school was made private in 2003, government schools have become over-crowded inducing some families to seek out schools with smaller class sizes (Bold et al. (2014); Oketch et al. (2010)). The results reported in Section 5 show that pupil-teacher ratios in Bridge are smaller than nearby government and private schools. This difference could be an important source of learning differences between Bridge and government schools if pupils face fewer distractions, are more comfortable during class, or can receive more individual attention.

Pedagogically, the Bridge model could differ in many ways from nearby schools, aside from its use of scripted lesson plans. Bridge has developed its own textbooks and learning materials (such as student homework books) and aims to provide each student with their own set of books.⁵⁸ This is in contrast with public schools where materials are often shared (Martin and Pimhidzai (2013); Uwezo (2016)).⁵⁹ If the quality and per-pupil availability of the Bridge textbooks differs from the alternatives in the market, then this could affect test scores and other outcomes.

Bridge students may also receive more or less homework assignments than students in government schools. Because Bridge students spend more time in class, they may spend less time completing homework.

The language of instruction in Bridge schools may also be different and this may mean that Bridge has different effects on different language skills. The SDI data show that many teachers use local languages in classroom observation. Bridge schools provide most instruction in English.

11.2 Heterogeneous effects

The results of tests of heterogeneous effects across different sub-groups defined by individual covariates will be reported. We will look at heterogeneity both to understand what mecha-

⁵⁸Some stakeholders argue that these materials may not be well aligned with the national curriculum.

⁵⁹It should be noted that, during this study government schools were providing additional course materials to Standard 1 and Standard 2 classrooms through the Tusome program.

nisms drive results and because some types of heterogeneity are of intrinsic interest for some policymakers.

- Grade Level - We will test the hypothesis that the model employed by Bridge is more suited to teaching lower-level material. For example, early grade instruction may require less content knowledge from teachers. This study will therefore look separately at effects for pre-primary (baby class, nursery, and pre-unit), lower primary (standards 1-4) and upper primary (standard 5-6).⁶⁰
- We will test the hypothesis that Bridge has a more positive or less negative impact when alternatives are weaker or for disadvantaged pupils by looking at heterogeneity by
 - Average KCPE score in county,
 - Average KCPE score of nearest government school, and
 - Household income per capita.
- We will also look for heterogeneity by
 - Number of years that a Bridge academy had been operating at the time of the scholarship
 - Rural/urban/peri-urban status - Classifications of academies into rural, urban, or peri-urban categories come from data supplied by Bridge.
 - Pupil Gender
 - Baseline academic achievement:

⁶⁰While we will test whether effects in each grade are statistically different from one another, it is important to note that these results are not dispositive evidence that Bridge is more or less effective in certain grades. An alternative explanation is that differences in treatment effects in the types of students who applied to the scholarship program at different grade levels could driving any observed heterogeneity in treatment effects between grade levels.

- * T3ET Math/English scores - because these scores represent results from multiple tests that vary across baseline schools and grades, the scores will first be transformed into percentile ranks within grade and geography cells.
- * Self-Regulation score (pupils enrolled through IPA only)
- * Malawi Development Assessment Tool (MDAT) (pupils enrolled through IPA only, in grades Baby class - STD 1)
- * Ravens Matrices (pupils enrolled through IPA only, STD 2 - STD 7)

Additional heterogeneity tests will be conducted to explore whether the effect of Bridge is more positive or less negative when teacher characteristics are indicative of greater levels of human capital. We will also test the triple interactions with the grade level of the applicant to explore the hypothesis that teacher human capital is more important in later grades where the material is more advanced. The characteristics used for these tests are described below.

- Teacher experience - Teacher experience data are available from the following sources:
 - Bridge administrative data can be used to identify the years of experience for the teacher instructing a given subject for a given grade. For teachers hired before 2016, data is not available on years of experience prior to employment at Bridge. After 2016, it is possible to identify total years of experience teaching for each teacher.
 - The 2016 local education environment survey includes data on the number of years of experience teaching *at the school where they were working at in 2016* for over 34,653 teachers. This data includes 1,337 Bridge teachers. These data will allow for the construction of variables that indicates the expected years of experience of a teacher outside of Bridge. It will then be possible to form a variable that expresses, for each academy, grade and subject combination, the expected difference in teacher experience between a teacher in a Bridge school and a teacher in a government school.

- Teacher content knowledge - Bridge administrative data include some subject knowledge scores recorded at training and hire. These records also include data on teachers' KCSE subject test scores.

11.3 Income effect of the scholarship

Under the assumption that winning the scholarship only affects pupil test scores through its effect on attendance 2SLS identifies the impact of attending Bridge rather than a combination of this effect and an income effect. Although this assumption is commonly invoked in the literature evaluating the effects of encouragement designs using vouchers, it may not be appropriate in some settings. The results of analyses described above will shed light on whether this assumption is reasonable in the case of this scholarship program. As discussed in Section 9.3, we will test if the scholarship influenced pupil test scores beyond its effect on attendance. If we find evidence of a significant direct effect of the scholarship we will try to develop adjusted estimates either using the decomposition methods described above in Section 9.3 or using a plug-in estimator described in Section 9.2.3.

Several tests, described above, will be examined in assessing the plausibility that the scholarship influenced pupil test scores beyond its effect on attendance.

11.3.1 Expenditures

The study will test the null hypothesis that the scholarship did not influence educational expenditures on students receiving the scholarship. Given that a substantial share of students attend private schools in the absence of the scholarship, the scholarship may have reduced direct fees paid by households. However, the scholarship may influence other education expenditures for several reasons. Households may redirect the money saved to other educational expenses or they may substitute away from or toward other education expenditures in response to any changes in educational inputs that the scholarship induced.

One channel through which the scholarship could influence students is by relaxing house-

hold budget constraints. By reducing the amount households had to spend on tuition, households could spend more on other inputs in the production of human capital. A common extra-curricular education service in this context is private tutoring, usually with local area teachers serving as the tutors. This option is available to many households in the study. The availability of this option motivates a test of the null hypothesis that households receiving the scholarship were not more likely to attend private tutoring sessions, attend more sessions, and have higher tutoring expenditures.⁶¹ Also, because such services are more often utilized by older children, the study will examine the effect of the scholarship on private tutoring services among younger and older students separately.

If an effect on other educational expenditures is observed, an additional interpretation is that parents believe non-school educational inputs are complements or substitutes with school inputs or with Bridge in particular. For example, an increase in non-school educational expenditures could come from parents of children who would have attended government or other private schools in the absence of the scholarship if those parents believe that private tutoring is complementary to Bridge.

11.3.2 Missed class

The study will test the null hypothesis that the scholarship did not influence the amount of class that pupils attended. The scholarship may have reduced the likelihood that pupils missed fee payments and were excluded from class if a pupil was likely to attend a private school in the absence of the scholarship. If missing class is harmful to test scores, one source of any treatment effect may be that treated students were less likely to miss class due to missed fee payments.

The effect of the scholarship on missed class will therefore be partially informative about whether the scholarship only affects pupils through their attendance at Bridge schools. If non-recipients miss more school than scholarship winners, then the scholarship winners will

⁶¹It is worth noting that the Bridge school day is longer than many other schools and pupils may simply not have as much time to attend extra tutoring sessions, even if the household is able to afford them.

simply have received more hours of instruction than non-winners and will have faced fewer disruptions.

These results will also be contextualized by other results regarding the amount of instructional time that pupils receive in Bridge and other schools. Bridge schools may, even after adjusting for missed class, provide more instructional hours if their school days are longer, if teachers are less likely to be absent, and if the school is more likely to have Saturday classes. Section 5 shows provides evidence that Bridge schools operate longer school days than most nearby schools (see Tables 3, 4, and 5 specifically). Subsequent analysis will compare the length of school day in different school types using pupil reports from the endline survey. Specifically, we will test the null hypothesis that the scholarship did not influence the pupil's length of school day and exposure to Saturday classes. The constellation of evidence will need to be considered in determining whether any observed effects on missed class could be driving observed effects on other pupil outcomes such as test scores.

11.3.3 Sibling spillovers

The results presented in Section 7.3 indicate that the scholarship may have induced siblings of scholarship winners to attend Bridge. Among the control group, a spillover effect from the treatment of siblings provides an opportunity to evaluate the effect of attending Bridge at cost. Siblings in the control group do not receive a scholarship when they attend Bridge, but if the experimental variation in treatment due to scholarship assignment among siblings induces these students to attend Bridge, then it is possible to estimate the effect of attending Bridge at cost.

Nevertheless, it is possible that the income effect of the scholarship itself also has a spillover effect on siblings who do not receive the scholarship. Even for a scholarship non-recipient, if many siblings win the scholarship, this may have an income effect on the household if some of those students would have required tuition payments in the absence of the scholarship. In addition to testing for effects on pupil test scores, the study will also test the

hypothesis that scholarships have no effect on aggregate household education expenditures, that expenditures on non-recipient siblings are the same regardless of sibling scholarship status, and whether non-recipient siblings are not more likely to miss class due to missed fees. These tests will shed light on whether the experimental variation in enrollment status induced by siblings can be informative about the effect of paying for Bridge. We will also examine whether non-recipient siblings who are induced to attend Bridge by their siblings' scholarships receive more tutoring sessions, as in Section 11.3.1.

11.4 Mechanisms and inputs in education production

This section discusses how results of the tests described above will be used to shed light on mechanisms driving any observed effects of Bridge on those who were induced to attend by the scholarship. This study will consider several mechanisms that may be particularly relevant in education production including class sizes, teacher absence rates, length of school day, or the use of technology to standardize lesson plans.

The randomized variation used for identification of effects in this study comes only from the scholarship program. The results of the study regarding the inputs driving the effect will therefore rely on additional assumptions and may only be suggestive.

11.4.1 Direct or proxy measurement of inputs

Using data from the endline survey, it is possible to estimate the effect of winning the scholarship on many variables that may be regarded as inputs in education production. Consider the stylized two sector model where there are Bridge schools B and government schools G . If outcome y is the output from a student achievement production function $f(X) = X\beta$, where X is a $K \times 1$ vector of inputs, and Bridge uses inputs X_B and government schools use inputs X_G , then the effect of moving a pupil from a government school to a Bridge school is represented by

$$\rho = (X_B - X_G)\beta = \Delta\beta .$$

If an input X_j is observed, Δ_j can be estimated using 2SLS. The reduced form is

$$X_{ij} = \alpha_0 + \alpha_1 Z_i + \epsilon_{ij} ,$$

where Z_i indicates whether the pupil received the scholarship and X_{ij} is the observed amount of input j reported by pupil at endline. We adopt this simple framework to interpret the results of several of the tests described above. For example, as described in Table 24, the study will estimate the effect of the scholarship on measures of overcrowding, parental interaction with school staff, access to textbooks, teacher responsiveness, teacher absences, and pupil effort. The study will also estimate the effects of the scholarship on the number of instructional hours per day using pupil reported school start and end times.

The interpretation of these results is complicated by the fact that, in most cases, the observed input j is a proxy for some input j^* . The results below will rely on parent and pupil reports about characteristics of the school, classroom, or teacher that may be correlated with j^* . The interpretation of the results must consider the possibility that reports on these surveys may not, in expectation, be consistent measures of the actual inputs received. Some measures are subjective assessments made by parents or pupils, and their responses may reflect expectations that are themselves influenced by the schooling environment. For example, a pupil who is exposed to frequent teacher absences may assess teacher absences differently than a pupil who is exposed to infrequent teacher absences. Results will therefore compare, when possible, pupil assessments with descriptive data from surveys of schools.

11.4.2 Heterogeneous effects

As described above, this study will examine heterogeneity of effects across a variety of covariates. This section describes in greater detail the interpretation of these effects when the covariates under study are interpreted as inputs in the education production process.

Variation in exposure to some inputs is observable within schools or between geographic

areas. This variation can be used to estimate the heterogeneous effects of treatment across different observed levels of the input.

Consider, again, the two sector model where each student i chooses between a local government school G and a Bridge school B in their local area a . The treatment effect for the school is then given by

$$\rho_a = (X_{Ba} - X_{Ga})\beta = \Delta_a\beta_a .$$

If Δ_{ja} (the difference in the input j between Bridge and government schools) is observed and it is not uniform across classrooms, it is natural to test whether the effect of Bridge is larger where the input j is larger. Specifically - assuming that all non-recipients attend government schools and all recipients attend Bridge schools for ease of exposition - for a endline outcome, it would be possible to estimate

$$y_{ai} = \alpha_0 + \alpha_1 Z_i + \alpha_2 \Delta_{aj} + \alpha_3 Z_i \times \Delta_{aj} + \epsilon_{ai} .$$

It is possible to identify the effect of Δ_{aj} , the difference between the input j in the Bridge school at the government school under the assumption that any correlation between Δ_{aj} and other subject varying inputs Δ_{aik} , $k \neq j$ is negligible. If Δ_{aj} measures the difference in the class size of Bridge and government schools, it is possible that this correlates with other characteristics of the schools, and interpretation will need to take into account other inputs that could be correlated with input j . If a separate measure of Δ_{aik} is available, it is possible to control for this in the specification.

In some cases, it may be possible to form measures of Δ_{ij} using data from the local education environment survey for those pupils who applied for scholarships at academies that were included in this survey activity.⁶² This survey collected data from all schools in

⁶²See Section 4.2.1 for details. The survey was only completed for the areas surrounding 78 academies, however some data are available from 47 additional academy areas. Our preferred specifications will either restrict to the 78 completed areas or restrict to the set of academy areas where at least one Bridge and

the area around the Bridge school making it possible to measure school characteristics in both Bridge and non-Bridge options available to students. This survey includes data on class size, teacher qualifications (whether they have a secondary degree, a teaching certificate, or diploma) and the years of experience of teachers. The data also includes data on the ages of pupils taught by each teacher in the school and their individual characteristics. This information allows for matching scholarship applicants to teachers within schools. The study will use these data to examine whether Bridge schools have a more positive or less negative effect on test scores when the nearest government school is has a higher pupil-teacher ratio, has a shorter school day, has more experienced teachers, and has more qualified teachers (as measured by the share of teachers with a teaching certificate or diploma). By restricting the comparison to Bridge schools and the nearest government school, it is hoped that any complications arising from the fact that the survey was not completed in some areas will pose less of a risk to identification.

In other cases, only X_{Bij} or X_{Gij} are observed, so additional assumptions are required about the distribution of X in the unobserved sectors is required. The 2017 survey of 54 randomly selected Bridge schools contains rich data on schools and teachers. Because the survey only includes Bridge schools, it is not possible to compare inputs used at Bridge schools to those at other schools in the area. When only the inputs at Bridge schools are observed, then it is possible to identify the effect of X_{Bij} only if it is assumed that the correlation between the residual variance in X_{Bij} and X_{Gij} , after controlling for other covariates is negligible.

Administrative data made available by Bridge on recruitment and training provide will be used to identify teacher qualifications, experience, and subject knowledge. Incoming staff are quizzed on several subjects, including mathematics, reading, and Kiswahili. Data is also collected on teacher's Kenyan Certification of Secondary Education (KCSE) exams. These data come from files that were updated in 2016, so teachers recruited after 2016 are not

one government school were surveyed. Restricting the sample in this way, the analysis using these data will contain 3,341 pupils total pupils.

included in these files.⁶³ These data can be used to estimate whether the effect of Bridge is larger when the pupil's would-be teacher has more experience, has more than one year of experience, has greater subject knowledge, or scored higher on the KCSE exam.

For all analysis of heterogeneous effects across measures of teacher characteristics, separate results will be reported for pupils in standard 4 and above, following the hypothesis that these students receive more advanced instruction and that teacher effectiveness in these grades may be more sensitive to the capabilities of the teacher.

11.5 Variation in effects across academies

The study will examine heterogeneity in effects of academies. Following Walters (2015), we will begin by testing for the presence of heterogeneity in effects across sites using a test of the goodness of fit of a model that predicts site effects entirely on the basis of the first stage effect of the scholarship on attendance. As noted in Walters, this test is equivalent to an over-identification test for 2SLS model with instruments formed by interactions between site dummies and treatment.

The results of this test may motivate additional follow-up analyses. Again, following Walters (2015), if there is evidence of heterogeneity in site effects, the standard deviation of site effects can be estimated using a two-sided selection model, avoiding problems that arise from estimating academy-specific effects using 2SLS in small samples.

These direct tests of the presence of heterogeneity in effects across sites will complement tests described above that will examine heterogeneity in effects across specific site-level covariates (rural/urban, average KCPE scores, age of academy).

These results will also be complemented by non-experimental estimates of the dispersion of academy value-added using Bridge's internal test data to estimate academy value-added. Academy value-added will be estimated using two separate non-experimental strategies. First estimates of the value-added of academies will be estimated as the effect of attend-

⁶³We are working with Bridge to try to obtain more recent files.

ing Bridge for longer. This is similar to the neighborhood exposure estimator in Chetty et al. (2014). Second, standard OLS value-added will be estimated for each school using a specification with controls, including a quartic expansion of prior period test scores. Variation in academy effects will be estimated following procedures from the teacher value-added literature (Rockoff (2004)) and neighborhood effects literature (Chetty and Hendren (2018)).

The results of these tests will be taken to be suggestive of the degree of standardization afforded by the Bridge model and, as such, may be informative about the scalability of the model (Cohodes et al. (2019)).

11.6 Effect on dispersion of test scores

This study will examine the effect of Bridge on the dispersion of test scores using multiple approaches.

First, we will test for equality of distributions using a Kolmogorov-Smirnov test.

Second, we will examine quantile specific effects using the Instrumental Variable Quantile Regression (IVQR) method of Chernozhukov and Hansen (2013). We will examine the effect of Bridge on the conditional quantiles at the bottom and top of the distribution of test scores.

Third, we will examine the heterogeneous effects of Bridge across measures of baseline achievement, as described above.

Fourth, we will directly compare the standard deviation of endline test scores in the treatment and control groups. The difference in the standard deviation between treatment and control groups provides a measure of the effect of the scholarship program itself on the dispersion of test scores, and is therefore partially informative about the effects of an voucher program on inequality of student test scores.

11.7 Heterogeneity across teacher skill levels

As described above, this study will test whether the effect of Bridge depends in part on characteristics of individual teachers. This section describes these tests in greater detail and

discusses interpretation of these results.

This study will test the hypothesis that the effect of Bridge is equal in classrooms led by teachers with varying credentials, experience, KCSE test scores, and subject knowledge scores. Several studies suggest that these characteristics may be important for classroom value-added. Bold et al. (2017) show that subject and pedagogical knowledge are predictive of within student variation between subject test scores. Outside of the developing world, teacher experience has been found to be an important input in teacher effects. The teacher value-added literature has found that teachers in their first years on the job are not as effective as they will be later in their careers. Kane et al. (2008), Chetty et al. (2014), and Staiger and Rockoff (2010) document that teachers in their first years tend to have effects of -0.08 to -0.5 SDs, but that they converge the mean effect rapidly over time.

This study will use data on teachers from the local education environment survey and administrative files provided by Bridge to test for heterogeneous effects across teachers with different levels of experience, credentials, and skills.

For each survey we will use an algorithm to match teachers to students.⁶⁴ The local education environment Survey collected data on grades and subjects taught by each teacher included in the survey to facilitate this matching process. These data, however, allow for multiple teachers to be associated with a single grade-subject pair. In these circumstances, we will take the average characteristic of teachers for the grade-subject pair. Matching administrative files on teachers to classrooms requires a crosswalk as the administrative personnel files do not record the grades and subjects taught by the teacher. The crosswalk is provided by the teacher value-added files maintained by Bridge. Because Bridge does not maintain files on value-added for pre-primary classrooms, this process cannot match teachers to pre-primary students. These files also occasionally identify multiple teachers of the same grade subject pair. In these cases, we will take the average characteristic of teachers identified with that grade-subject pair.

⁶⁴Error from mismatching students to teachers will induce measurement error in the covariate, and this will attenuate effects toward zero.

Given that the intervention took place over two academic years, a second component of the matching process is to determine what teachers to match students to. Our preferred approach will match students to their 2017 teacher (or, in the event that they are being matched to a prospective teacher, to the teacher who would have taught them had they attended a particular type of school). This choice is motivated in part because the effects of characteristics of 2016 teacher may have partially faded out by the end of the 2017. Results will be examined using 2016 teachers and using the average of characteristics of teachers in 2016 and 2017 classrooms.

After matching students to their prospective Bridge teacher for each subject, we will first test the hypothesis that the effect of Bridge is unaffected by observable characteristics of the teacher. This test will be repeated for the aggregate subject learning index and for each subject individually. Specifically, where characteristic X_{jcs} is the characteristic of the teacher for subject s at academy j for class c , we will estimate the following structural model using 2SLS.

$$Y_{is} = \beta_0 + \beta_1 D_i + \beta_2 X_{jcs} + \beta_3 X_{jcs} \times D_i + \epsilon_{is}$$

Interpretation of the effect measured by the coefficient β_3 is complicated by the fact that teacher characteristics may be correlated with other unobserved characteristics of the education market or Bridge academy. In order to address this complication, an additional test will be performed that tests specifically the hypothesis that Bridge is equally effective when subjects are taught by teachers with different levels of each characteristic measured relative to the market or academic mean.

$$Y_{is} = \beta_0 + \beta_1 D_i + \beta_2 X_{jcs} + \beta_3 X_{jcs} \times D_i + \beta_4 \bar{X}_j + \beta_5 \bar{X}_j \times D_i + \epsilon_{is}$$

Where \bar{X}_j is an estimated of the expectation of X within the education market indexed by j .

A related test comes from Bold et al. (2017). Pooling all subject tests, including student fixed effects, and restricting to the sample of students who actually attended Bridge in 2017, the hypothesis will be tested that characteristic X_{jcs} predicts Y_{is} , the pupil’s performance on the subject s assessment.

Finally, in order to describe the extent to which individual teachers are driving the estimated effect of Bridge, we will first measure teacher value-added on academic subject achievement using historical test scores from Bridge classrooms. Following the literature on the estimation of teacher value-added models.⁶⁵ Treating these measures of value-added in a manner analogous to the discussion of other characteristics, we will estimate the effect of being instructed by a teachers at different locations in the teacher value-added distribution. These results will allow for comparisons of the effectiveness of Bridge when being instructed by some of the worst/best teachers at Bridge schools.

11.8 Heterogeneity across class sizes

As documented in Section 5, Bridge schools tend to have lower class sizes than government schools and many nearby private schools. If a positive effect of Bridge is identified, it is possible that the effect is attributable to the small class sizes that Bridge provides. Understanding the role of class size in producing any observed positive effects of Bridge is important for using the results of this study to extrapolate to other settings where class size will not be small. For example, if Bridge continues to grow, some of this growth will likely come by increasing enrollment within their currently operating schools. Importantly, this study pro-

⁶⁵By the phrase “teacher value-added”, we are referring to the a technique commonly used in the social science literature (see for instance Kane et al. (2008); Rockoff (2004); Chetty et al. (2014)). Teacher value-added models are used to estimate the causal effect of a teacher or classroom j on pupil i ’s test score in time period t , y_{it} . Writing the pupil’s test score as $y_{it} = X'_{it}\beta + \mu_j + \epsilon_{it}$, where X_{it} are potentially time varying observable characteristics of pupil i , μ_j is a teacher (or classroom) effect, and ϵ_{it} is an idiosyncratic error, we can try to estimate μ_j using fixed effect estimation if we observe a panel of student test scores for pupil i . Chetty et al. (2014) find that such measures can be reliable measures of the causal effect of teachers on student outcomes, especially when X_{it} contains a pupil’s test score from an earlier period, $y_{i,t-1}$. In addition to a quartic polynomial expansion of $y_{i,t-1}$, we will include in the controls the pupil’s gender, their age, and a quartic expansion of the number of days that have elapsed since the pupil enrolled in Bridge, and interactions between the baseline score $y_{i,t-1}$ and the number of days that the pupil has been enrolled in Bridge.

vides estimates of the effect of Bridge that are internally valid for a setting where Bridge’s classes are relatively small. Extrapolating from the results of this study to answer questions about the effects of Bridge with substantially larger classrooms requires assumptions about the effects of increasing class size. Evidence from rich countries finds that class size may have an effect on student test scores. Schanzenbach (2006) reviews the results from the Tennessee Star experiment showing that the small class experiment has significant effects on student test scores. Angrist and Lavy (1999), using a regression discontinuity design, find effects of smaller class size on test scores for students in 4th and 5th grade, but not for younger students in 3rd grade. The evidence from developing countries seems to indicate that class size may not have a large impact on pupil test scores. Duflo et al. (2015) find that reducing class size in Standard 1 classrooms in Kenya had no effect on test scores. In that study, class size was reduced from 82 to 44 on average as a result of the intervention. It is possible that at more moderate class sizes, class size may be more or less important in the production of student achievement. However, Holla and Kremer (2009) reviews three other studies from developing countries, including one from Kenya, that have identified the effects of class size using randomized variation for classrooms with range of baseline class sizes.⁶⁶

To begin to address this question, we will test whether the effect is more positive or less negative in classrooms with fewer pupils. For this test we will estimate class size using Bridge’s administrative test score data file. We will use class size in 2017 as the endogenous class size. To avoid including the pupil’s own endogenous attendance decision in the interaction term, we will use the projected size of the classroom using Bridge’s 2015 test data file as an instrument. The result of this test must be interpreted with caution. Large class size indicate greater demand, and greater demand may be correlated with other characteristics of the Bridge academy. For example, if low class size indicate a low quality Bridge academy,

⁶⁶It is possible that, in the context of Bridge schools, class size may matter more if it is the case that other school resources are used more efficiently or teachers are less likely to be absent. For example, Duflo et al. (2015) found that some teachers may have exerted lower effort in response to smaller teacher class size intervention as evidenced by higher rates of teacher absences. If the Bridge model effectively monitors teachers so that teachers are already exerting a high level of effort, then teachers may not be able to adjust their level of effort as class size changes. If that is the case, class size could matter more in Bridge schools.

the test described above might find that the effect of Bridge is more positive or less negative when class size is larger, even if the effect of class size is zero or negative.

We will extend this specification first by including an additional interaction for the average class size at the school. This strategy will rely more on within school variation in class size which may be less influenced by unobserved differences in quality between Bridge schools and nearby schools. We will also examine the relationship between class size and student outcomes using Bridge’s administrative test data. For example, we will test whether class size is associated with lower pupil performance controlling for academy fixed effects.

11.9 Vulnerability to manipulation and external validity

Because the scholarship status of pupils was publicly observable, it is possible that *ex post* manipulation by some party could have influenced the results. One concern is that scholarship winners were given special treatment and therefore the estimated effects for these students, while internally valid, lack external validity.

The research team will explore the possibility of extra support for scholarship winners and the closely related question of whether the results from the scholarship evaluation have external validity for students who might pay out-of-pocket for a Bridge education. The results presented on the effects of Bridge at cost, discussed above, are related to this question. If Bridge has a large effect on those attending on a scholarship, but a smaller effect on those attending at cost, this is suggestive that Bridge may have given scholarship winners special treatment.⁶⁷ Additionally, we will compare the estimated treatment effects to observational, non-experimental methods. These will include a comparison of observational value-added estimates of Bridge separately for students attending Bridge on scholarship and at cost. Specifically, we will estimate OLS value-added of Bridge using baseline controls separately for scholarship winners and losers.

⁶⁷Although it is important to note that this result could be explained by either a direct effect of the scholarship on outcomes or treatment effect heterogeneity due to the fact that those who attend Bridge at cost may be different than those who attend when the cost is zero.

11.10 Cost Effectiveness

The estimated effects described above will be used as inputs in cost-benefit analysis. The research team will not use this document to describe in detail our plans for this analysis because pre-specifying such plans does not reduce the scope for unconscious or conscious bias. The research team has already seen some data on costs at both Bridge and government schools and will continue to seek out accurate data on costs at Bridge schools and other nearby schools. These data will be used to conduct cost-benefit analyses under a variety of assumptions. The results of these analyses will be presented transparently so that any assumptions being made are clearly noted and so that interested readers can consider the effects of relaxing those assumptions.

References

- Abadie, A., Chingos, M. M., and West, M. R. (2018). Endogenous stratification in randomized experiments. *The Review of Economics and Statistics*, 100(4).
- Abdulkadirođlu, A., Angrist, J. D., Hull, P. D., and Pathak, P. A. (2016). Charters without lotteries: Testing takeovers in new orleans and boston. *American Economic Review*, 106(7):1878–1920. Link.
- Abdulkadirođlu, A., Pathak, P. A., Schellenberg, J. T., and Walters, C. R. (2019). Do parents value school effectiveness. *NBER Working Paper*.
- Abdulkadirođlu, A., Pathak, P. A., and Walters, C. R. (2018). Free to choose: Can school choice reduce student achievement. *American Economic Journal: Applied Economics*.
- Andrews, I., Stock, J., and Sun, L. (2019). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*.
- Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *American Economic Review*, 92(4):1535–1558.
- Angrist, J. and Lavy, V. (1999). Using maimonidies’ rule to estimate the effect of class size on student achievement. *Quarterly Journal of Economics*.
- Behrman, J. R. and Parker, S. W. (2013). Is health of the aging improved by conditional cash transfer programs? evidence from mexico. *Demography*, 50(4). Link.
- Benhassine, N., Devoto, F., Duflo, E., Dupas, P., and Pouliquen, V. (2015). Turning and shove into a nudge? a “labeled cash transfer” for education. *American Economic Journal: Economic Policy*, 7(3):86–125. Link.

- Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., and Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in africa. *Journal of Economic Perspectives*, 31(4).
- Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2013). The high return to private schooling in a low-income country. Technical report, Africa Growth Initiative at Brookings. Link.
- Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2014). Can free provision reduce demand for public services? evidence from kenyan education. *World Bank Economic Review*, 29(2):293–326. Link.
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(439).
- Burde, D. and Linden, L. L. (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*.
- Chernozhukov, V. and Hansen, C. (2013). Quantile models with endogeneity. *Annual Review of Economics*.
- Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–79.
- Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.
- Cohodes, S., Setren, E., and Walters, C. R. (2019). Can successful schools replicate? scaling up boston’s charter school sector. *NBER Working Paper*.

- Dean, J. T. and Jayachandran, S. (2019). Attending kindergarten improves cognitive but not socioemotional development in india. Unpublished.
- Duflo, E., Dupas, P., and Kremer, M. (2015). School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from kenyan primary schools. *Journal of Public Economics*, 123:92–110. [Link](#).
- Gibbs, C., Ludwig, J., and Miller, D. (2013). Does head start do any lasting good? In Bailey, M. J. and Danziger, S., editors, *Legacies of the War on Poverty*, pages 39–65. Russell Sage Foundation Press, New York.
- Haushofer, J. and Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from kenya. *The Quarterly Journal of Economics*. [Link](#).
- Holla, A. and Kremer, M. (2009). Improving education in the developing world: What have we learned from randomized evaluations? *Annual Review of Economics*. [Link](#).
- Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from chile’s voucher program. *Journal of Public Economics*, 90:1477–1503.
- Hull, P. (2018). Isolating: Identifying counterfactual-specific treatment effects with cross-stratum comparisons. Manuscript.
- Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review*, 27(6):615–631.
- Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3).

- Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: the case of head start. *The Quarterly Journal of Economics*, 141(4):1795–1848.
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.
- Kolen, M. J. and Brennan, R. L. (2014). *Test Equating, Scaling, and Linking*. Springer-Verlag, New York, New York.
- Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *Review of Economics and Statistics*.
- Martin, G. H. and Pimhidzai, O. (2013). Service delivery indicators: Kenya. Technical report, The World Bank. [Link](#).
- Muralidharan, K. and Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *The Quarterly Journal of Economics*, 130(3):1011–1066.
- Oketch, M., Mutisya, M., Ngware, M., and Ezech, A. C. (2010). Why are there proportionately more poor pupils enrolled in non-state schools in urban kenya in spite of fpe policy. *International Journal of Educational Development*, 30(1):23–32. [Link](#).
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economics*.
- Piper, B., Zuilkowski, S. S., and Mugenda, A. (2014). Improving reading outcomes in kenya: First-year effects of the primr initiative. *International Journal of Educational Development*.
- Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., and Downer, J. (2012). Tthird grade follow-up to the head start impact study final report. Technical

- report, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Link.
- Reardon, S. F. and Raudenbush, S. W. (2013). Under what assumptions do site-by-treatment instruments identify average causal effects. *Sociological Methods & Research*, 42(2):143–163.
- Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review (Papers and Proceedings)*.
- Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the milwaukee parental choice program. *The Quarterly Journal of Economics*, 113(2):553–602.
- Schanzenbach, D. W. (2006). What have researchers learned from project star. *Brookings Papers on Education Policy*.
- Staiger, D. O. and Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, 24(3):97–118.
- Stocking, M. L. and Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*.
- Uwezo (2016). Are our children learning? uwezo kenya sixth learning assessment report. Technical report, Nairobi: Twaweza East Africa. Link.
- Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from head start. *American Economic Journal: Applied Economics*, 7(4):76–102.
- Witte, J. F. (1998). The milwaukee voucher experiment. *Educational Evaluation and Policy Analysis*.

A Randomization Details

This section describes, in detail, the procedures used to assign scholarships to students. These procedures include basic data cleaning to remove duplicates, the application of eligibility requirements, stratification, and the institutional constraints that determined the number of scholarships to be awarded at each academy.

The randomization proceeded in two rounds.

A.1 Round 1

29,969 scholarship applications were received for students in grades baby class to STD 7. Bridge also received applications for students in STD 8, but these applicants were assigned by Bridge.⁶⁸

Prior to the first round of randomization, 3,667 were removed in total because they already received scholarships through other programs administered by Bridge internally, because they did not meet eligibility requirements, or because they were duplicates.

20 applications were found to be duplicates. 111 applications were found to be from students who were awarded scholarships through other programs. Two applications were dropped because no academy was indicated.

UnitedWeReach (UWR) required that, for students applying to standard 4 and above, applicants needed to have a score of 40 percent or above on their term 3 end-of-term exam. 944 applicants did not meet this requirement. UWR also assigned a composite score, combining socioeconomic and academic factors for students who were already attending Bridge. These factors are described in Appendix A.3. Students with an index below one were ineligible for the scholarship.

The selection score was also used to assign some scholarships through a deterministic

⁶⁸It is the research team's understanding that Standard 8 applicants received scholarships through a deterministic process administered by Bridge. In total 539 Standard 8 applicants were given scholarships.

process. At each academy, the student with the highest selection score was awarded a scholarship. 404 scholarships were assigned through this process. These applicants were removed from the sample prior to randomization.⁶⁹

The number of seats available to be randomly assigned in each strata was determined through a process that tried to ensure fair distribution and to avoid overcrowding. Several constraints were established in advance of the first round of randomization by stakeholders involved in the program.

- The total number of scholarships at an academy could not exceed 15 percent of projected enrollment.
- The total number of scholarships at an academy for current Bridge pupils could not exceed 35 percent of 15 percent of projected enrollment.
- The total number of scholarships at an academy for children not currently enrolled in Bridge could not exceed 85 percent of 15 percent of projected enrollment.

Note that in each case, the total number of scholarships allowed *includes* scholarships assigned prior to the randomization process. Bridge had awarded 1,168 scholarships to current Bridge students in grades below standard 8, 539 scholarships to standard 8 students, and 110 scholarships to students who were not currently Bridge students.

In addition to these constraints, the number of scholarships within a strata was restricted so that the offer rate did not exceed 60 percent.

Some strata contained only a single applicant. These strata were treated separately. These applicants are referred to as “small cell” applicants. Applicants in strata with two or more applicants are referred to as “large cell” applicants.

⁶⁹One academy, Sichirai (Kakamega county) did not have any applicants who were currently enrolled at the academy, so no student was awarded the scholarship at this academy.

Table A1: Sample breakdown prior to randomization

Likely counterfactual	Large cell	Small cell	Deterministic
<i>Panel A: New to Bridge in 2015</i>			
Government school	11,650	91	0
Private school	2,300	141	0
<i>Panel B: At Bridge in 2015</i>			
Bridge	11,656	60	404

Prior to randomization students in large cells were grouped according to the student's expected school choice in the absence of the scholarship.

- S1 : Students who were not attending Bridge and expected to attend government schools in large cell strata
- SS1 : The same, but in small cell strata
- S2 : Students who were currently enrolled in Bridge schools
- SS2 : The same, but in small cell strata
- S3 : Students who were not attending Bridge and expected to attend private schools
- SS3 : the same but in small cell strata

The total number of seats available for students in the S1 sample is then given by

$$Q_a^{S1} = \min \left\{ 0.6 \times N_a^{S1}, \min \left\{ 1 - 0.15 - \frac{L_a}{E_a}, 1 - \frac{D_a}{E_a} \right\} \times (0.15 \times E_a) \right\}$$

Where N_a^{S1} is the total number of S1 applicants to academy a , E_a is the projected enrollment at academy a , D_a is the number of scholarships assigned to current Bridge applicants prior to randomization through other programs, and L_a be the number of new to Bridge applicants assigned scholarships prior to randomization through other programs.

This resulted in 5,183 available scholarships for S1 students.

The number of seats available for S2 applicants was then determined by the following:

$$Q_a^{S2} = \min \{0.6 \times N_a^{S2}, \max \{ \min \{0.35 \times 0.15 \times E_a, 0.15 \times E_a - Q_a^{S1} - L_a - D_a\} - D_a \}, 0 \}$$

This resulted in 2,932 scholarships for $S2$ students.

The number of seats available for SS2 applicants was then determined by the following

$$Q_a^{SS2} = \min \{N_a^{SS2}, \max \{ \min \{0.35 \times 0.15 \times E_a - (Q_a^{S2} + D_a), 0.15 \times E_a - (Q_a^{S1} + Q_a^{S2} + D_a + L_a)\}, 0 \} \}$$

In other words, if, after assigning scholarships to $S1$ Q_a^{S1} , Q_a^{S2} and assigning scholarships through deterministic to current Bridge students D_a and new students L_a , neither the 15 percent of projected enrollment nor the 15 percent of 35 percent of total enrollment thresholds had been crossed, then scholarships would be made available up to the number of SS2 N_a^{SS2} applicants.

This resulted in 29 scholarships for SS2 students.

An analogous rule was applied for determining the number of SS2 applicants.

$$Q_a^{SS1} = \min \{N_a^{SS1}, \max \{ \min \{0.85 \times 0.15 \times E_a - (Q_a^{S1} + L_a), 0.15 \times E_a - (Q_a^{S1} + Q_a^{S2} + Q_a^{SS2} + D_a + L_a)\}, 0 \} \}$$

This resulted in 29 scholarships for SS1 students.

Again, the an analogous rule was applied to determine the number of SS3 applicants.

$$Q_a^{SS3} = \min \{N_a^{SS3}, \max \{ \min \{0.85 \times 0.15 \times E_a - (Q_a^{S1} + Q_a^{SS1} + L_a), \\ 0.15 \times E_a - (Q_a^{S1} + Q_a^{S2} + Q_a^{SS1} + Q_a^{SS2} + D_a + L_a)\}, 0 \} \}$$

This resulted in 111 scholarships for SS3 students.

Finally, the number of scholarships available for S3 students was determined by the same rule.

$$Q_a^{S3} = \min\{N_a^{S3}, \max\{\min\{0.85 \times 0.15 \times E_a - (Q_a^{S1} + Q_a^{SS1} + Q_a^{S3} + L_a), \\ 0.15 \times E_a - (Q_a^{S1} + Q_a^{S2} + Q_a^{SS1} + Q_a^{SS2} + Q_a^{SS3} + D_a + L_a)\}, 0\}\}$$

This resulted in 690 scholarships available for S3 students.

A.2 Round 2

A second round of randomization was required after it was recognized that some students would not utilize the scholarship and that more than 10,000 scholarship would need to be offered in order to allocate 10,000 scholarships. Also, after the first round, it was discovered that some students had received scholarships through other programs administered by Bridge directly. Note that scholarship winners from the first round *had not* been informed of their status prior to the second round of randomization.

Prior to the second round of randomization 5 duplicates were identified, all of whom were in the control group of the S2 sample.

The second round of randomization offered new scholarships to only S1 and S2 applicants.

In the second round, the maximum number of scholarships allowed at an academy was increased from $\frac{3}{20} \times E_a$ to $\frac{E_a}{3}$. However, if this number exceeded 50 percent of the number of S1 applicants to academy a , N_a^{S1} , then no more scholarships were allowed for S1 applicants.

A similar rule was employed for the S2 applicants, but in this case, only 8 percent of target enrollment was allowed to go to scholarships.

The second round of scholarships allocated 769 scholarships to S1 applicants and 1,172

S2 applicants. No additional scholarships were issued to S3 applicants in this round.

A.3 Selection Criteria

Applicants to the scholarship program who were already in Bridge schools were assigned a selection score which was used to determine eligibility for the scholarship program, as described in Appendix A.1. The overall score was a composite of three sub-scores:

- Sibling Factor (35 percent)
- Academic Achievement Factor (32.5 percent)
- Vulnerability Factor (32.5 percent)

A.3.1 Sibling Factor

Students received 1 point for every sibling attending Bridge.

A.3.2 Academic Achievement Factor

The academic achievement factor was assigned using information either on the pupil's rank or term 3 end of term test score.

Table A2 describes how points were assigned.

Table A2: Academic Achievement Index

Point assignment	Class rank	Term 3 test score
3	1	> 450
2	2-3	400-449
1	4-5	350-399
0.5		300-349
0		< 300

Notes: Students were either assigned their academic achievement factor according to their class rank or their Term 3 end of term test score. If an applicant had both a class rank and a T3ET test score, they received the maximum points associated with either their rank or their test score.

After forming the academic selection index, the academic selection factor was formed by assigning points for each quintile of the academic selection index:

- 1st quintile: 0 points
- 2nd quintile: 1 point
- 3rd quintile: 2 points
- 4th quintile: 3 points
- 5th quintile: 4 points

A.3.3 Vulnerability Factor

The vulnerability factor was formed by first constructing a vulnerability index. The index was constructed using the following table.

The applicant’s vulnerability factor was calculated as the applicant’s quintile in the vulnerability index distribution minus one.

B Endline Sampling

This section describes the sampling procedure used to select applicants for endline follow-up.

As described in the text, all PP_{govt} and P_{govt} students were selected for endline.

Approximately 22 percent of the PP_{brig} and P_{brig} students was selected for the endline survey. Sampling of these applicants was performed by selecting clusters of applicants by the original randomization strata⁷⁰. Prior to sampling, the randomization strata cells (the sampled unit) were stratified into 5 bins defined by quintiles of the distribution of the number of total applicants in the cells. 22 percent of units within each quintile bin were the selected for follow-up. Most strata contained more controls than treated individuals. Within each strata s , only $\min\{N_{ts} + 2, N_{cs}\}$ control units were selected for follow-up, where N_{ts} denote the number of treated individuals in unit s .

⁷⁰The strata formed when initially offering the scholarships, described in Appendix A

Question - Short	Point Assignment	Category
Pupil disability, specifically mobility?	Yes - 1	Pupil Disability
Mother alive?	No - 1 Don't Know - 0.5	Orphanhood
Father alive?	No - 1 Don't Know - 0.5	
Caretaker 1 can read?	No - 1	Caretaker Education
Caretaker 2 can read?	No - 1	
Caretaker 1 can write?	No - 1	
Caretaker 2 can write?	No - 1	
Caretaker 1 is deaf?	Yes - 1	Caretaker Disability
Caretaker 1 is blind?	Yes - 1	
Caretaker 1 has mobility issues?	Yes - 1	
Caretaker 2 is deaf?	Yes - 1	
Caretaker 2 is blind?	Yes - 1	
Caretaker 2 has mobility issues?	Yes - 1	
# of Adults in Household	Average Per Capita Resource: Total Income / (# Adults + # Children) > 50th Percentile - 0 26-50th Percentile - 1	Resource Availability
# of Children in Household	10-25th Percentile - 2	
Total Household Income	< 10th Percentile - 3	
Total # of Income Earners in Household	>1 - 0 1-Jan 0 - 0	
Rent or own?	Rent - 1	Expenditures
Has electricity?	No - 1	Assets
Has a latrine?	No - 1	
Material of floor of home?	Earth - TK Wood - TK Cement - TK Tiles - TK	
Material of walls of home?	Cane or Planks - TK Mud - TK Mabati - TK Timber - TK Bricks - TK Cement - TK	
Asset ownership	Own Sim - 0 Mobile Phone - 0 Radio - (-0.5) Television - (-0.5) Video/DVD player (-0.5) Computer - (-2) Sewing Machine - (-1) Gas cooking stove - (-1) Electric cooking stove - (-2) Refrigerator - (-2) Bicycle - (-1) Motor cycle - (-3) Car - (-5) Non - 1	

All P_{priv} applicants were included in the endline. However, the procedure for allocating scholarships gave this group of applicants the lowest priority. In many cases, this procedure resulted in many randomization strata where applicants had no risk of being assigned a scholarship. These applicants would not contribute to the experimental variation in treatment assignment being used for estimation, and therefore would not increase the precision of the result. Therefore, these students were excluded from endline activities. 841 P_{priv} applicants in total were found in randomization strata cells that had trivial (either 0 or 1) risk of assignment to treatment.

C Assessment Development

The research team contracted a team of consultants to develop a series of assessments (32 individual tests) for this evaluation. The consultants were well suited to this task as they they had extensive teaching experience in Kenya –combined the team had over 70 years of teaching experience in Kenya. The consultants were all TSC trained teachers who had experience in both government and private schools in Kenya. They also had experience working with students from similar socioeconomic backgrounds as the participants in our study. Finally, one of the consultants had over 20 years of experience working at the Kenya Institute for Curriculum development. The consultants were supported by research assistants.

The consultants developed 32 specific tests covering math, Swahili, English, and science/social studies (combined) for grades 1 through 8. To ensure that the test items (or questions) were aligned with Kenyan curriculum, the team started with the published curriculum (printed in 2012), and a research assistant outlined the key skills in each subject-grade to provide a summary to aid the test development process. The consultants also used copies of mid-term and end-of-year tests gathered from Kenyan public schools, and a variety of KICD approved test prep books, including those published by Targerter, to strengthen the curricular alignment and ensure that the tests were similar in format, style, and content

to those that students in Kenyan schools would typically encounter.

The consultants developed a bank of questions and ranked these questions as “easy”, “medium”, and “hard”. Consultants submitted 3 questions of each skill difficulty for each subject-grade. The items in the test banks were then edited and reviewed by the research team to ensure adequate balance in the question bank. This was an important step to ensure that the test avoided floor or ceiling effects. The team also tried to ensure that the questions contained a balance of “rote” and “critical thinking” questions. The balance in the tests were evaluated by the research assistants who mapped the questions to the skills outlined in Blooms taxonomy.

For the ECE tests, the research team combined items from several international tests including MELQO, UWEZO, EGRA/EGMA, and other items from Kenyan test prep materials and previous ECE evaluations conducted in Kenya.

To select the final test items, the draft versions of the tests were piloted several times with students who were from similar socio-economic backgrounds as our sample. The results of these pilots were evaluated using classical test theory (e.g. Cronbach’s alpha). Questions that were too easy, too difficult, or were uncorrelated with remaining items (item-rest correlation) were either revised or removed. The pilots also yielded information on the amount of time required to complete each test question. Based on the results of this pilot a final assessment was developed for each grade.

All tests included items that overlapped with tests in adjacent grades to facilitate the development of an IRT measure. To facilitate bench-marking to national and international tests, the tests also included materials from TIMMS/ PIRLS and Uwezo.

Table C1 presents results on the internal reliability of each assessment using Cronbach’s α .

Table C1: Assessment reliability: Cronbach's α

	English		Kiswahili		Math		Science and Social Studies	
	α (1)	# Items (2)	α (3)	# Items (4)	α (5)	# Items (6)	α (7)	# Items (8)
Pre-primary	0.93	32	0.93	22	0.95	45		
Standard 1	0.93	32	0.95	36	0.86	18	0.71	15
Standard 2	0.92	29	0.92	29	0.89	20	0.64	18
Standard 3	0.90	27	0.91	26	0.86	19	0.72	17
Standard 4	0.83	23	0.85	23	0.70	19	0.73	16
Standard 5	0.85	22	0.84	23	0.77	21	0.74	15
Standard 6	0.82	18	0.78	23	0.80	21	0.75	18
Standard 7	0.78	18	0.67	20	0.76	19	0.70	17
Standard 8	0.72	20	0.62	21	0.79	20	0.73	19

Notes: The table reports Cronbach's α estimates for each assessment. Cronbach's alpha is calculated in the following way. For a test with K items, the sum of correct responses is denoted by Y and x_k is an indicator for whether the item k was answered correctly. Cronbach's α is calculated as $\frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_k^2}{\sigma_Y^2} \right)$, where σ_k^2 is the variance in scores on item k , and σ_Y is the variance in scores on Y .

D Formal LATE decomposition

This appendix will illustrate the multiple counterfactuals/mediators problem extending the model in Hull 2018 to allow for 4 sub-LATEs. In order to ensure that the discussion is comprehensive, the exposition considers the problem of fully decomposing the voucher effect into each of the sub-LATEs. This includes the possibility that the voucher had a direct effect on “always takers” (those students who would have enrolled in Bridge regardless of treatment status) because it reduced the amount households had to pay.

Consider the case of estimating the LATE for the effect of going to Bridge *for free* for any subsample of students participating in the scholarship program. Let $Z_i \in \{0, 1\}$ indicate whether student i received the scholarship. Students choose an educational environment $j \in \{B, G, P, U\}$, where $B \in \{0, 1\}$ indicates *paying* to go to Bridge⁷¹, $G \in \{0, 1\}$ government school, $P \in \{0, 1\}$ a non-Bridge private school, and $U \in \{0, 1\}$ non-enrollment/home care. Let j_i indicate attendance at a school of type j , j_{zi} indicate the potential outcome given treatment $Z_i = z$. Note that $B_{1i} = 0$ for all i .

Write the potential outcome of attending Bridge on scholarship as Y_{1i} . Potential outcomes in all of the other states $j \in \{B, G, P, U\}$ will be written as Y_{ji} . The observed outcome can be written (suppressing the subscript i) as

$$Y = Y_1 + (Y_B - Y_1)B + (Y_G - Y_1)G + (Y_P - Y_1)P + (Y_U - Y_1)U$$

Under standard assumptions for two stage least squares (independence, exclusion, and monotonicity), generalized to account for the multiple counterfactuals⁷² it is possible to derive the following convenient expression for the difference in outcomes between treatment and control.

⁷¹In this example, all of the endogenous variables will be treated as binary variables, although the results are similar if the endogenous variable is allowed to take on multiple values.

⁷²These assumptions are described formally in Appendix E

$$\begin{aligned}
E[Y|Z = 1] - E[Y|Z = 0] &= E[Y_1 - Y_B|B_1 < B_0]Pr(B_1 < B_0) \\
&+ E[Y_1 - Y_G|G_1 < G_0]Pr(G_1 < G_0) \\
&+ E[Y_1 - Y_P|P_1 < P_0]Pr(P_1 < P_0) \\
&+ E[Y_1 - Y_U|U_1 < U_0]Pr(U_1 < U_0) \quad (5)
\end{aligned}$$

This expression is derived in appendix E.

It is worth emphasizing that $Pr(j_1 < j_0)$ is the first stage effect of treatment on attendance at a school of type j . The first stage regression of D (Attending Bridge on scholarship) on Z is

$$\begin{aligned}
\pi_D &= E[D|Z = 1] - E[D|Z = 0] \\
&= Pr(B_1 < B_0) + Pr(G_1 < G_0) + Pr(P_1 < P_0) + Pr(U_1 < U_0) \quad (6)
\end{aligned}$$

Writing the first stage effect on the remaining choices of educational environment j , $Pr(j_1 < j_0|X)$, as $\pi_j(X)$ it is possible to write a “decomposed” expression for the LATE estimated by using the scholarship to estimate the effect of attending Bridge on scholarship:

$$\begin{aligned}
LATE &= E[Y_1 - Y_B|B_1 < B_0] \frac{\pi_B}{\pi_D} \\
&+ E[Y_1 - Y_G|G_1 < G_0] \frac{\pi_G}{\pi_D} \\
&+ E[Y_1 - Y_P|P_1 < P_0] \frac{\pi_P}{\pi_D} \\
&+ E[Y_1 - Y_U|U_1 < U_0] \frac{\pi_U}{\pi_D} \quad (7)
\end{aligned}$$

Equation 7 shows that the LATE estimated for the effect of attending Bridge *for free* is

a weighted average of sub-LATEs describing the relative effectiveness of *Bridge for free* and the four counterfactual educational environments.

The empirical objective is to identify the relative effectiveness of each of 4 counterfactuals. The first step is to form a $k \times 1$ multivariate function of $p \times 1$ column vector of baseline characteristics X_i , $M(X_i)$, where $k \geq 3$. Multiple instruments are then formed by interacting $M(X_i)$ with treatment. With $k = 3$, there are 4 excluded instruments ($[Z_i M(X_i)]'$ $[Z_i]$) so the estimating equation is exactly identified.

The heterogeneous effect of treatment on each of the endogenous variables across applicants with different values of $M(X_i)$ is used to identify the separate effects associated with each sub-LATE. Identification requires the additional maintained assumption that variation in the LATE across students with different values of $M(X_i)$ be negligible. Following Hull (2008), this assumption can be stated formally as follows

Assumption D.1 (LATE homogeneity) $E[Y_{1i} - Y_{ji} | j_{1i} < j_{0i}, M(X_i)] \approx E[Y_{1i} - Y_{ji} | j_{1i} < j_{0i}, M(X'_i)]$ for any X_i and X'_i .

As described above, the existing literature has made use of many functions $M()$. This function could select a subset of covariates from X_i or it could be the product of prediction algorithms.

E Derivation of LATE decomposition expression

This appendix derives the expression obtained for the LATE decomposition into multiple counterfactuals. The proof follows Hull (2018)'s proof of Lemma 1.

As in the text, the subscript i is dropped. Note that the derivation is the same if the object of interest is the LATE conditional on some stratifying variables X_i , as in the case of Hull (2018).

Following Hull (2018), the following assumptions are made:

Assumption E.1 (Independence) *Potential outcomes are independent of treatment status*

Assumption E.2 (Exclusion) $E[Y_{0j} = Y_{1j} \forall j \in \{1, G_i, B_i, P_i, U_i\}]$

Assumption E.3 (Monotonicity) $Pr(j_1 \leq j_0) = Pr(k_1 \leq k_0) = 1$

$$E[Y|Z = 1] - E[Y|Z = 0] = E[Y_1 + (Y_B - Y_1)B + (Y_G - Y_1)G + (Y_P - Y_1)P + (Y_U - Y_1)U | Z = 1] \\ - E[Y_1 + (Y_B - Y_1)B + (Y_G - Y_1)G + (Y_P - Y_1)P + (Y_U - Y_1)U | Z = 0]$$

$$= E[Y_1 | Z = 1] - E[Y_1 | Z = 0] \\ + E[(Y_B - Y_1)B_1 | Z = 1] - E[(Y_B - Y_1)B_0 | Z = 0] \\ + E[(Y_G - Y_1)G_1 | Z = 1] - E[(Y_G - Y_1)G_0 | Z = 0] \\ + E[(Y_P - Y_1)P_1 | Z = 1] - E[(Y_P - Y_1)P_0 | Z = 0] \\ + E[(Y_U - Y_1)U_1 | Z = 1] - E[(Y_U - Y_1)U_0 | Z = 0]$$

Applying the independence assumption, this can be written as

$$= E[(Y_B - Y_1)(B_1 - B_0)] \\ + E[(Y_G - Y_1)(G_1 - G_0)] \\ + E[(Y_P - Y_1)(P_1 - P_0)] \\ + E[(Y_U - Y_1)(U_1 - U_0)]$$

Applying the monotonicity assumption, this can be written as

$$\begin{aligned}
&= E[Y_1 - Y_B|B_1 < B_0]Pr(B_1 < B_0) \\
&\quad + E[Y_1 - Y_G|G_1 < G_0]Pr(G_1 < G_0) \\
&\quad + E[Y_1 - Y_P|P_1 < P_0]Pr(P_1 < P_0) \\
&\quad + E[Y_1 - Y_U|U_1 < U_0]Pr(U_1 < U_0)
\end{aligned}$$

F Leave-one-out counterfactual predictions

This section describes an approach to predicting counterfactual educational choices using academy level means. This approach is closely related to the approach to endogenous stratification described by Abadie et al. (2018) in that it interacts treatment with a predicted outcome using observed data for the control group. That study shows that predictions that leave individual i in individual i 's own predicted outcome risk overfitting to the control group. Comparisons of treatment effects in the treatment and control groups of an experiment can therefore be biased due to mean reversion arising from the fact that extreme predictions in the control group are unlikely to be accurate in the treatment group. That study finds that leaving individual i out of individual i 's own prediction reduces the risk of overfitting to the control group. In the scholarship evaluation, the objects of prediction are probabilities of the probability that an individual i attends either a government school, private school, or Bridge school, and that a student is unenrolled.

Variable	Interpretation
$j \in \{G, B, P, U\}$	School Environment choice (endogenous variable)
$t \in \{2016, 2017\}$	Year of school environment choice
G_{it}	Attendance at government school in year t
B_{it}	Attendance at Bridge <i>at cost</i> in year t
P_{it}	Attendance at non-Bridge Private school in year t
U_{it}	Un-enrolled in year t
s	Academy \times Sample index

F.1 Leave-one-out academy means

The object being predicted is the likelihood that a pupil would have attended education option j in the absence of the scholarship. This is estimated using a leave-one-out mean of each of the attendance variables. The means are formed separately for each of the subsamples (PP_{gout} , P_{gout} , PP_{brig} , P_{brig} , and P_{priv}) separately.

Because many of these cells are small, the predictions may be imprecise. A method to improve the precision of the resulting 2SLS estimators is to shrink imprecise predictions based on small samples toward the overall mean. The next section describes this process.

F.2 Empirical Bayes Shrinkage: Binomial with Beta prior

This section describes an approach to estimating the empirical Bayes posterior prediction by treating the likelihood distribution of a choice in each year t as a binomial and the prior as a Beta distribution.⁷³

The academy means are constructed for each year of the experiment $t \in \{2016, 2017\}$ at the academy level. Each endogenous attendance variable is treated separately⁷⁴, and the mean for the cell is assumed to be drawn from a binomial distribution characterized by a parameter θ_{jts} . An empirical Bayes posterior is then computed that shrinks the raw means toward the overall mean for the sample. A Beta distribution, $B(\alpha_{jt}, \beta_{jt})$, is estimated using maximum likelihood for each year t and education option j .

To simplify the exposition, we drop the subscripts j , t , and s and describe the method for computing the empirical Bayes posterior for a single cell.

With a parameter θ , the probability of observing n out of k individuals in education option j is given by

$$p(k, n, \theta) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} .$$

⁷³The Beta distribution is chosen as a conjugate prior for the binomial distribution. The posterior of a binomial with a Beta prior is a Beta distribution.

⁷⁴The object of prediction is a multinomial distribution. In future work, we may consider incorporating the multinomial structure into the empirical Bayes prediction.

To form the empirical Bayes posterior we first estimate the prior hyperparameters of the Beta prior, α and β , using maximum likelihood. These estimates are formed using the distribution of cell mean attendance at school type j across cells s .

With a binomial likelihood function and a Beta prior we have

$$p(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{p(x)},$$

where x is observed data on n and k . Plugging in the distribution of the binomial and beta distributions, we get

$$p(\theta|x) = \left(\binom{n}{k} \theta^k (1-\theta)^{n-k} \right) \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \right),$$

where $\Gamma()$ represents the gamma function. This expression can be rearranged⁷⁵ to show that the posterior is a $B(\alpha+k, \beta+n-k)$. The posterior mean is then

$$\hat{\mu}^{post} = \frac{k + \hat{\alpha}}{n + \hat{\alpha} + \hat{\beta}}.$$

This procedure is repeated for all sample cells s , for each school type j , and for teach year $t \in \{2016, 2017\}$.

G Construction of endogenous attendance variables

Data on school attendance were collected at several points during the study, including during the phone call surveys in 2016 and 2017 and also the endline survey. It is therefore possible to construct a variety of endogenous school attendance variables for which the receipt of the scholarship is a valid instrument. This section describes how the survey data are used to con-

⁷⁵To see this, note that

$$p(\theta|x) \propto \theta^{\alpha+k-1} (1-\theta)^{\beta+n-k-1}.$$

struct a variable indicating exposure to Bridge schools and other educational environments. First, the procedure for identifying whether a pupil was enrolled in Bridge in 2017, the final year of the scholarship program, is described. Second, the procedure for identifying the number of years the pupil was enrolled in Bridge between January 2016 and December 2017 is described. Finally, analogous procedures for exposure to other educational environments (government schools, non-Bridge private schools, non-enrollment) are described.

A complicating factor for IV estimates is that the endline survey was conducted over a relatively long period that spanned two academic years, from November 2017 to March 2018. Students may be more likely to switch schools at the start of the new academic year in January. Therefore, using information on the school attended by the student at the time they are interviewed may lead to mis-estimation of the first stage effect on school attendance. With this in mind, the survey collected information on the school attended during the 2017 academic year to ensure that attendance is measured uniformly across students.

It is possible that attendance in 2017 does not fully capture the degree of exposure of pupils to each education environment when the duration of the study was 2 years. For example, when exposure is defined in terms of 2017 attendance, pupils who dropped out Bridge at the close of the 2016 school year are treated the same as pupils who never set foot in a Bridge class room. An alternative measure that captures differences in “dosage” is the number of years exposed to each education environment. This document will describe approaches to calculating both using available data.

The parent survey collected information on school attendance of the pupil. The pupil survey collected similar data, but because some pupils may not be old enough to provide reliable information on the type of school that they are attending, parent responses are used first in creating the attendance data. Pupils were never asked to classify their school as being either a government school, Bridge school, or non-Bridge private school.

The caregiver survey was intended to capture information on whether pupils were unenrolled for long periods during the period under study. Caregivers who responded negatively

when asked whether their child was enrolled in school at the time of the survey were asked a follow-up question to ascertain whether the child was enrolled at any point in 2017. Due to a coding error, the survey skipped this item. 30 percent (4,167) of the endline data was collected using a form that included this error. As a result, 2,000 observations do not have data from the caregiver survey regarding whether the child attended any school in 2017.

The next two sections describe specific decisions that have been made to classify attendance for each pupil in 2016 and 2017.

G.1 2017 attendance classification

This subsection describes the procedure used to determine the type of school attended in 2017 by each pupil with endline data.

All pupils are classified as having attended Bridge in 2017 according to the caregiver’s response to a series of questions regarding the type of school that the pupil was attending in 2017. Additionally, if the caregiver indicated that the pupil was not currently enrolled in school and the pupil survey does not contradict this information (either the pupil doesn’t know or the item is missing), the pupil is classified as being unenrolled. Any pupil who reports that they were not enrolled in 2017 and have not been classified in a previous step are classified as unenrolled.

These procedures leave 95 pupils unclassified in 2017.

Next, for any pupil who reports that the name of their school⁷⁶ contains the word “bridge” and who had not been classified by any previous step, we classified the pupil as having been enrolled in Bridge. This step classifies 56 pupils as having attended Bridge in 2017.

Any pupil who was found to be attending a Bridge academy in the 2017 phone call survey and who is not classified by the procedures above is classified as having attended a Bridge academy in 2017. This step classifies 5 pupils as having attended Bridge in 2017.

⁷⁶This item asks the pupil what the name of the school is that they attended “in the last term of 2017”. For pupils who were unsure what time period was being referred to, the FO would prompt that this was around the time of the recent August elections.

A small number of pupils are classified on the basis of the 2016 phone call survey. Pupils who responded to the 2016 phone call survey and who reported in the endline survey that they had not switched schools since 2016 were classified according to their response in the 2016 survey.

The remaining pupils are classified based on the name of the school that the pupil reports in the endline survey. Only 19 pupils are classified using this procedure. Schools are classified as being government schools if the name of the school includes the phrase “primary” and includes a reference to the physical location of the school. Schools are classified as “private” otherwise. This decision rule was adopted by observing that the word “primary” is highly associated with government school attendance. 91 percent of government school attendees reported the word “primary” in their schools’ names, compared to whereas 9 percent of non-government school attendees.

G.2 2016 attendance classification

For pupils whose caregivers report that the pupil has been enrolled in their 2017 school for 2 or more years, the pupil’s 2016 school of attendance is classified as the same as the 2017 attendance. 2,688 pupils are unclassified after this first step.

For pupils who reported switching school between May 2016 and November 2017, the caregiver survey asked what type of school the pupil switched to and from. An additional 1,743 students are classified according to the school from which they switched.

The pupil survey included a question regarding the class in which the pupil was enrolled. An allowable response was “Never enrolled in school”. 22 children are classified as having been unenrolled in 2016 because this field indicates that they were never enrolled in school.

Of the 943 pupils who are not classified after the procedures described above, 698 pupils are classified by using reported attendance in the 2016 phone call survey directly. An additional 10 pupils are classified supplementing with the 2017 phone call survey data. Specifically, if a pupil reports not having switched schools since 2016 in the 2017 phone call survey,

they are classified as attending the same type of school that they were attending in 2017.

Of the 235 pupils, 195 pupils are classified as having attended their 2017 school in 2016 if the parent indicates that the pupils did not switch schools between May 2016 and November 2017 (the start of endline activities).

An additional 9 pupils are classified as unenrolled in 2016 based on field officer notes indicating that the pupil had never been enrolled in school.

The remaining 31 pupils are classified as having attended their 2017 school in 2016. It should be noted that this assumption that students who cannot be classified based on existing data have not switched school types will tend to make measure of the number of years of exposure collinear with 2017 attendance. The research team will examine the sensitivity of the results to treating these 31 pupils as missing data for the endogenous attendance variable. Given the small number students, it is unlikely that this choice will have a significant effect on the results.

G.3 2018 and 2019 attendance classification

For some outcomes collected in the 2019 tracking survey activities, it will be useful to be able to identify the type of school attended in 2018 and 2019. The 2019 phone call survey activity included data collection on KCPE test taking status and the KCPE score.

The 2019 phone call survey collected information on school attended in 2019. August 13, 2019, items were added to capture information on school type attended in 2018. Because this item was added late, it is missing for over 4,385 pupils (approximately 40 percent of completed surveys as of November 2019).

It is straightforward to identify school attended in 2019 using these data. Classification of school attended in 2018 is complicated by the fact that this information is missing for a large number of pupils.

All pupils who were called back to collect information on 2019 KCPE outcomes in December 2019 were asked for both 2018 and 2019 school attendance information. Therefore,

for this cohort of KCPE test takers, we anticipate full coverage attendance data for both 2018 and 2019. There still remain pupils who took the KCPE test in 2018 for whom information will not be available on 2018 school attendance status. To deal with these pupils, we will use an imputation procedure. Specifically, we will first estimate, using multinomial logit, the probability of attending each school type in 2018 given school type attended in 2017 and 2019. This will yield a prediction for each pupil, given their observed 2017 and 2019 attendance. This vector of predictions will then be treated as the parameters of a multinomial probability from which an imputed attendance vector will be drawn for each pupil with missing 2018 attendance data.

H Appendix Figures

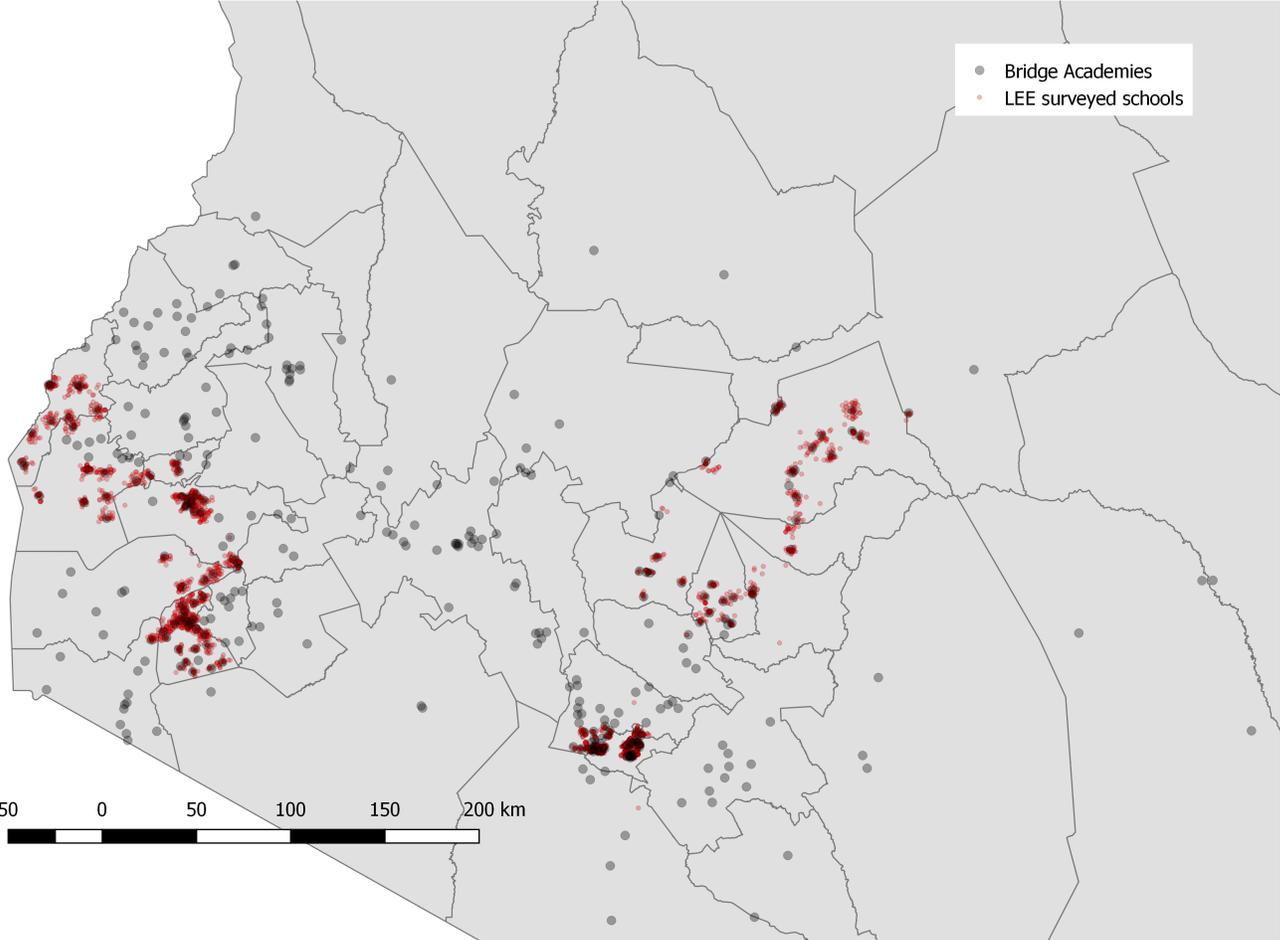


Figure 2: Map of schools surveyed in the Local Education Environment Survey.

I Appendix Tables

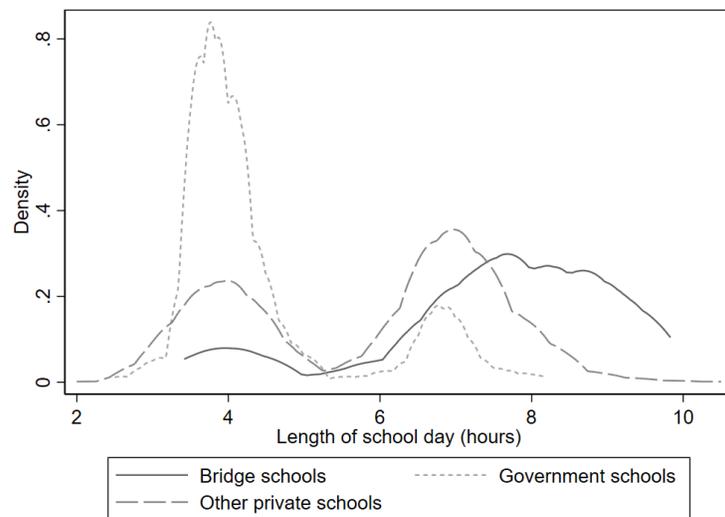


Figure 3: Distribution of pre-primary school day length by school type.

Table G1: Follow-up and differential attrition - Phone call tracking surveys

Sample	Full Sample			Baseline contact sample			IPA sample		
	Mean (1)	Diff. (2)	<i>N</i> (3)	Mean (4)	Diff. (5)	<i>N</i> (6)	Mean (7)	Diff. (8)	<i>N</i> (9)
<i>Panel A: 2016 Phone Call Survey</i>									
PP _{govt}	0.778	0.076*** (0.012)	4,423	0.799	0.069*** (0.014)	3,107	0.767	0.110*** (0.031)	802
P _{govt}	0.841	0.024*** (0.010)	5,286	0.862	0.021* (0.012)	3,541	0.874	0.023 (0.021)	1,061
PP _{govt} + P _{govt}	0.812	0.048*** (0.008)	9,709	0.833	0.044*** (0.009)	6,648	0.828	0.062*** (0.018)	1,863
<i>Panel B: 2017 Phone Call Survey</i>									
PP _{govt}	0.666	-0.003 (0.013)	4,423	0.681	-0.021 (0.016)	3,107	0.645	-0.040 (0.034)	802
P _{govt}	0.714	0.025** 0.012	5,286	0.732	0.018 0.015	3,541	0.736	0.048* 0.028	1,061
PP _{govt} + P _{govt}	0.692	0.012 0.009	9,709	0.708	-0.001 0.011	6,648	0.697	0.009 0.022	1,863

Notes: Standard errors are reported in parentheses. Each coefficient estimate of differential attrition comes from a regression of an indicator for endline test score follow-up for the sample listed on the left hand column. The prefix ‘P’ refers to primary school, and the prefix ‘PP’ refers to pre-primary. The subscript *govt* refer to pupils who were determined, using baseline data from the application, to be likely to attend government schools in the absence of a scholarship. Only PP_{govt} and P_{govt} pupils were included in the 2016 phone call survey. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table G2: Test equating coefficients using Stocking-Lord method

		English				Math				Kiswahili				Science and Social Studies			
		$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c	$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c	$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c	$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
ECD	Est.	-	-	1.40	-2.17	-	-	1.34	-4.19	-	-	1.20	-1.94	-	-	-	-
	SE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
STD 1	Est.	0.81	1.18	1.13	-1.21	1.16	2.26	1.55	-1.58	0.76	1.31	0.91	-0.94	-	-	0.26	-7.96
	SE	(0.03)	(0.03)	-	-	(0.06)	(0.07)	-	-	(0.03)	(0.03)	-	-	-	-	-	-
STD 2	Est.	0.90	0.71	1.02	-0.70	0.92	0.96	1.42	-0.56	1.10	0.54	1.00	-0.49	1.48	1.07	0.39	-6.38
	SE	(0.03)	(0.04)	-	-	(0.04)	(0.05)	-	-	(0.03)	(0.05)	-	-	(0.09)	(0.09)	-	-
STD 3	Est.	1.13	0.79	1.15	-0.05	0.85	0.51	1.20	-0.10	1.01	0.47	1.02	-0.09	1.83	1.42	0.71	-2.52
	SE	(0.03)	(0.05)	-	-	(0.04)	(0.04)	-	-	(0.04)	(0.05)	-	-	(0.19)	(0.17)	-	-
STD 4	Est.	0.87	0.07	1.00	0.00	0.83	0.13	1.00	0.00	0.98	0.10	1.00	0.00	1.40	0.66	1.00	0.00
	SE	(0.04)	(0.05)	-	-	(0.05)	(0.05)	-	-	(0.05)	(0.06)	-	-	(0.09)	(0.09)	-	-
STD 5	Est.	0.98	0.42	0.98	0.29	1.12	0.65	1.12	0.55	1.18	0.71	1.18	0.70	1.22	0.46	1.22	2.12
	SE	(0.04)	(0.05)	-	-	(0.06)	(0.06)	-	-	(0.06)	(0.07)	-	-	(0.08)	(0.07)	-	-
STD 6	Est.	0.96	0.15	0.94	0.40	1.27	0.53	1.42	1.11	1.11	0.38	1.30	1.10	1.16	0.44	1.41	4.52
	SE	(0.04)	(0.05)	-	-	(0.06)	(0.06)	-	-	(0.05)	(0.06)	-	-	(0.07)	(0.07)	-	-
STD 7	Est.	1.02	0.33	0.97	0.62	1.09	0.65	1.55	1.86	0.99	0.25	1.28	1.37	1.22	0.85	1.73	10.12
	SE	(0.06)	(0.07)	-	-	(0.06)	(0.06)	-	-	(0.05)	(0.06)	-	-	(0.08)	(0.08)	-	-
STD 8	Est.	1.02	0.53	0.99	1.00	1.25	0.74	1.94	2.92	1.05	0.68	1.34	2.13	1.00	0.22	1.72	11.57
	SE	(0.08)	(0.09)	-	-	(0.09)	(0.08)	-	-	(0.07)	(0.08)	-	-	(0.08)	(0.08)	-	-

Notes: This table presents the equating coefficients using the Stocking-Lord method between assessment levels for each subject test.

Table G3: Test equating coefficients using Stocking-Lord method

		English				Math				Kiswahili				Science and Social Studies			
		$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c	$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c	$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c	$x_{c-1,c}$	$y_{c-1,c}$	x_c	y_c
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
ECD	Est.	-	-	0.98	-2.59	-	-	0.94	-3.07	-	-	0.83	-2.46	-	-	-	-
STD 1	SE	0.97	1.39	0.95	-1.20	0.97	1.52	0.91	-1.55	1.17	1.31	0.97	-1.15	-	-	0.47	-1.14
STD 2	Est.	0.96	0.63	0.91	-0.59	1.00	0.94	0.90	-0.65	1.00	0.52	0.97	-0.55	1.06	0.45	0.50	-0.69
STD 3	SE	1.10	0.50	1.00	-0.12	0.95	0.52	0.86	-0.15	0.99	0.40	0.96	-0.08	1.49	0.37	0.74	-0.30
STD 4	Est.	1.00	0.12	1.00	0.00	1.16	0.16	1.00	0.00	1.05	0.07	1.00	0.00	1.35	0.19	1.00	0.00
STD 5	SE	0.99	0.36	0.99	0.37	1.20	0.44	1.20	0.47	1.08	0.39	1.08	0.47	1.18	0.20	1.18	0.44
STD 6	Est.	1.00	0.15	0.99	0.51	1.21	0.37	1.46	0.94	1.04	0.25	1.13	0.80	1.15	0.22	1.36	0.98
STD 7	SE	1.02	0.21	1.02	0.73	1.10	0.53	1.60	1.76	1.00	0.20	1.13	1.06	1.11	0.46	1.51	2.33
STD 8	Est.	1.10	0.27	1.11	1.01	1.20	0.49	1.92	2.59	0.97	0.51	1.09	1.75	1.18	0.24	1.78	3.10

Notes: This table presents the equating coefficients using the concurrent estimation method described in Section 10.1.1.

Table G4: Item numbers to be included in the local content knowledge score

Standard	Local tools & foods (1)	Health & Infectious disease (2)	Civics & culture (3)	Geography (4)	Industry (5)
3	3,5	11	14,16		7
4	6,4	1	16		12,13
5	14	4,5		12	10,11,13
6		3	15,16,17,18	14,17	12,13
7		5	16	14,15	12,13,17
8		5	18,19	17	14,15

Notes: This table shows the item numbers from each assessment that will be included in the local content knowledge scores. Each row represents an assessment level for a particular Standard. The item number refers to the item number in the science and social studies section of the assessment for the standard indicated by the row labels.