WORKING PAPER · NO. 2022-68

Can Education be Standardized? Evidence from Kenya

Guthrie Gray-Lobe, Anthony Keats, Michael Kremer, Isaac Mbiti, and Owen Ozier

JUNE 2022

Development Innovation Lab at University of Chicago

AFFILIATE OF Becker Friedman Institute

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Can Education be Standardized? Evidence from Kenya

Guthrie Gray-Lobe* Anthony Keats† Michael Kremer‡ Isaac Mbiti§
Owen Ozier ¶

Working Paper

This version: June 6, 2022

We examine the impact of enrolling in schools that employ a highly-standardized approach to education, using random variation from a large nationwide scholarship program. Bridge International Academies not only delivers highly detailed lesson guides to teachers using tablet computers, it also standardizes systems for daily teacher monitoring and feedback, school construction, and financial management. At the time of the study, Bridge operated over 400 private schools serving more than 100,000 pupils. It hired teachers with less formal education and experience than public school teachers, paid them less, and had more working hours per week. Enrolling at Bridge for two years increased test scores by 0.89 additional equivalent years of schooling (EYS) for primary school pupils and by 1.48 EYS for pre-primary pupils. These effects are in the 99th percentile of effects found for at-scale programs studied in a recent survey. Enrolling at Bridge reduced both dispersion in test scores and grade repetition. Test score results do not seem to be driven by rote memorization or by income effects of the scholarship.

The evaluation received support from the World Bank Strategic Impact Evaluation Fund (SIEF), the World Bank Knowledge for Change Program (KCP), the World Bank Research Support Budget (RSB), the World Bank Education Global Practice (GEDDR), JPAL Post-Primary Education Initiative, the Bill and Melinda Gates Foundation, and the Omidyar Network's education initiative, which spun out to become Imaginable Futures in 2020. The scholarship program was funded by UnitedWeReach. We are grateful to Amrita Ahuja, Isaiah Andrews, Joost de Laat, Joshua Dean, David Evans, Ivan Fernandez-Val, Eric Hanushek, Susannah Hares, Peter Hickman, Andrew Ho, Alaka Holla, Seema Jayachandran, Heeyon Kim, Dev Patel, Ben Piper, Mauricio Romero, Juan Saavedra, and Justin Sandefur for helpful comments. We are grateful to William Blackmon, Maria Bucciarelli, Celeste Carano, Peter Hawes, Mridul Joshi, Naomi Kimani, Carol Nekesa and Ben Wekesa for excellent research assistance. We acknowledge the hard work and dedication of our field teams, and the support of Asman Suleiman and the IPA-Kenya management team. 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 obtained. Institutional Review Board (IRB)/Ethics approval was received from Maseno University and IPA. The scholarship evaluation was registered with the American Economic Association's registry for randomized controlled trials (AEARCTR-0005382). The registration includes a Preliminary Analysis and Plan For Subsequent Analysis (PAPSA) document that was developed while partially blinded to the data. We thank Joel Ferguson and the Berkeley Initiative for Transparency in the Social Sciences (BITSS) for conducting a code replication of the tables in the paper. The World Bank International Finance Corporation (IFC), Bill Gates, and the Omidyar Network invested in Bridge. Some authors of this study are conducting other research with Bridge.

*University of Chicago, email: graylobe@uchicago.edu
†Wesleyan University, email: akeats@wesleyan.edu
‡University of Chicago, email: kremermr@uchicago.edu
§University of Virginia, email: imbiti@virginia.edu
¶Williams College, email: owen.ozier@williams.edu

1 Introduction

> Our biggest challenge is that we need to ensure we standardize everything. If we want to be able to operate like McDonald's we need to make sure that we systematize every process, every tool – everything we do. – *Jay Kimmelman, Bridge co-founder, quoted in Rangan and Lee (2010)*

Standardization has spread across sectors from automobile manufacturing (e.g. Ford motors) to airlines (e.g. checklists for pilots) to restaurants (e.g. McDonald's). Developing and refining standardized processes requires some fixed costs and often adopting new technology, such as the assembly line in automobile manufacturing. However, it offers the promise of consistent product quality at large scale by using a few highly-skilled workers to provide detailed instructions to many frontline workers who actually manufacture products or deliver services. It often involves a host of complementary management changes, such as introduction of standardized monitoring systems and new human resource policies. Although standardization is sometimes used with highly educated and experienced workers, as for example with checklists for pilots and surgeons, it often involves hiring less educated and experienced frontline workers at lower salaries. As illustrated by Ford's introduction of standardized automobile production, the creation of the United Auto Workers as an industrial (rather than craft) union, and the eventual adoption of the National Labor Relations Act of 1935, standardization can sometimes generate social conflict and the eventual adoption of new social institutions to influence how standardization is used and how any productivity gains from standardization are shared among different stakeholders.

A priori, it is unclear whether education is a conducive setting for extensive standardization. Many argue that teachers need to draw on a range of techniques to best serve the individual learner, so a standardized approach is not appropriate (Dresser, 2012; Valencia et al., 2006; Lamb-Sinclair, 2017; Nunnery, 2021). Effective instruction may require tacit knowledge, including complex social interactions, which cannot be encoded in instructions (Polanyi, 1967). Some other attempts to produce standardized educational materials at large scale, such as MOOCs (Massive Online Open Courses), failed to fulfill initial optimistic expectations, arguably because they did not manage to motivate most students (Banerjee and Duflo, 2014).

Private business models based on standardizing operations at scale in low- and middle-income countries may also encounter other challenges. A literature in development economics notes the relative scarcity of large firms in low- and middle-income countries and argues that difficulties faced by productive private firms inhibit overall productivity growth. Hsieh and Klenow (2009) argue that large private firms are subject to tax and regulatory "wedges" that generate misallocation, while small informal firms operate under the radar screen and large state-owned enterprises are subsidized. Others attribute the scarcity of large firms in low income countries to weak legal structures (Bertrand and Schoar, 2006) and difficulties associated with maintaining control across multiple establishment locations when this requires managers outside the family (Kremer et al., 2019; Ilias, 2006).

We study a highly structured and standardized approach to education, implemented by Bridge International Academies (henceforth "Bridge") in Kenya. The Bridge operating model features centrally-developed, highly-detailed lesson plans (equivalently, "scripts" or "lesson guides") that are delivered to teachers using tablet computers. Supervisory staff (including headteachers) are trained and closely monitored to promote the use of lesson plans by teachers. In addition to pedagogical scripts for teachers, monitoring scripts provide a detailed framework for headteachers on how to conduct classroom observations, provide feedback, and thus increase teacher accountability. Standardized processes are also employed for other business operations, ranging from school construction to financial management.

At the time of the study, Bridge operated over 400 schools spread throughout most of Kenya's 47 counties, with an enrollment of over 100,000 pupils, accounting for approximately 1 percent of primary enrollments and 0.65 percent of pre-primary enrollments in the country.

We estimate the impact of enrolling in Bridge schools by using the random variation created by a large, nationwide program that offered 10,000 two-year scholarships to attend Bridge schools in any of the ten grade levels. With more than 25,000 applicants, the program was oversubscribed, and scholarships were awarded by lottery. These scholarships increased the likelihood that lottery winners enrolled in Bridge by about one-third.

We find that enrolling at Bridge improves student learning as measured by five subject knowledge tests designed to reflect the material in the national curriculum. Primary school pupils in our main sample (who were considered likely to enroll at a public primary school if they did not win a

scholarship),[1] gained 2.89 equivalent years of Kenyan schooling (EYS) after being enrolled at Bridge for two years, an additional 0.89 years compared to pupils enrolled at other schools over the same period.[2] This differential is equivalent to 0.81 standard deviations of the 4th-grade distribution of these tests. Pre-primary pupils gained 3.48 equivalent years of schooling (EYS), an additional 1.48 years compared to pupils enrolled at other schools. This differential is equivalent to 1.35 standard deviations on these tests. The test score effects in this study are among the largest observed in the international education literature, particularly for a program that was already operating at scale, exceeding the 99th percentile of treatment effects of large-scale education interventions reviewed by Evans and Yuan (2020).

We test for, and find no evidence that, the test score results are explained by Bridge emphasizing memorization at the cost of higher order skills. We find positive and statistically significant effects on higher order skills.[3]

Enrolling at Bridge reduces dispersion in test scores, having larger effects for pupils at the bottom of the test score distribution than those at the top. We do not see evidence that the effect of Bridge on test scores varies by other dimensions we examined, such as pupils' gender or socio-economics status, or the pupil-teacher ratio (PTR) in Bridge classrooms.

Beyond the reduced dispersion of test scores, we find additional evidence consistent with the idea that Bridge's efforts to standardize education produces relatively uniform outcomes, including across diverse teacher characteristics and locations. The effect of Bridge is similar even among teachers with less than three years of experience and low academic content knowledge, two characteristics that previously were found to be relevant determinants of teacher effects (Chetty et al., 2014; Bau and Das, 2021). In addition, the variance in effects across Bridge locations is small relative to the overall effect, and we are unable to reject the null hypothesis that the effect of Bridge is constant across individual schools (or sites). A model of selection that yields direct estimates of the standard deviation of Bridge effects on test scores across sites suggests that among primary school pupils, school mean test scores for Bridge pupils vary about 28 percent less than those outside

[1] As described in Section 3.3, we identify pupils who were likely to enroll at public primary school as those who were not enrolled at a Bridge school at the time of application and who either attended a public primary school at baseline (2015) or stated that they were planning to enroll at a public primary school in the first year of the scholarship program (2016). Section 5.1 confirms that these applicants predominantly enrolled in public schools.

[2] Subsection 3.5 discusses reasons for expressing effects in these units.

[3] Our measure of higher-order skills uses the same subject knowledge assessments but excludes items that potentially capture rote learning or recalling information.

Bridge. Among pre-primary pupils, the variation is larger by 29 percent for Bridge pupils than those outside Bridge.

Beyond academic test scores, we find that enrolling at Bridge has positive effects on grade progression and on measures of working memory, self-control, and receptive vocabulary. Enrolling at Bridge increases the probability that pupils advanced two grades after two years by 20.3 percentage points in primary school and 17.8 percentage points in pre-primary grades. To address concerns that the highly-structured pedagogy approach used by Bridge may detract from creativity (Riep and Machacek, 2016), we administered a test of divergent thinking. We find no statistically significant differences between Bridge scholarship recipients and other pupils.

We also looked at differences in various intermediate schooling inputs between pupils who were induced to enroll in Bridge and their counterparts who enrolled in other (primarily public) schools. Bridge schools have higher reported instructional time, lower reported teacher absence, and higher reported teacher and parent engagement in child's education.

Several pieces of data suggest that income effects do not drive the test score effects we observe and that it is reasonable to follow the literature in treating income effects of scholarships on test scores as negligible (e.g Dean and Jayachandran, 2019; Duflo et al., 2021). First, the scholarship did not increase expenditure on tutoring, a common non-school educational expenditure. Second, we compare the scholarship effect across sub-samples of the data for which the relative effects of the scholarships on Bridge enrollment and educational expenditures vary. We find that baseline characteristics are predictive of variation both in the scholarship's impact on educational expenditures and in Bridge enrollment.[4] We show that variation in the scholarship impact on test scores is explained by the pattern of enrollment effects and is largely invariant to the pattern of educational expenditure effects. We formally estimate the separate effect of educational expenditures and Bridge enrollment and discuss the assumptions required to disentangle these two effects. Because the results rely on stronger assumptions, they are not dispositive evidence against any income effects. However the evidence weighs strongly against the notion of large positive income effects driving the impact of the scholarship on test scores, because there are no large changes in expenditure patterns, and there is no evidence for income effects on learning in sub-samples (or

[4]These baseline characteristics include geographic location, enrollment in Bridge at baseline, the school type where applicants said they would enroll without the scholarship, and indicators of socioeconomic status.

populations) where income effects were strongest.

At the primary school level, our preferred estimate suggests that EYS gains from switching from other private schools to Bridge were comparable to EYS gains from switching from public schools to Bridge. However, an alternative estimator, subject to weak instrument concerns yields an estimate close to zero because public schools were the counterfactual option for the majority of pupils in our sample. Estimating the impact on test scores of enrolling at Bridge instead of other private schools requires imposing some maintained assumptions and yields noisy estimates. At the pre-primary school level, estimated effects of enrolling at Bridge relative to other private schools range from 1.10 to 1.80 EYS, but are subject to weak instrument concerns.

This study is not intended as a comprehensive evaluation of Bridge, and judgments about Bridge will depend on issues beyond those that can be analyzed with the data in this study. We can only bring limited evidence to bear on questions that have been raised about Bridge's compliance with the national curriculum, working conditions and pay for Bridge teachers, registration of schools, and student safety in schools (Education International and KNUT, 2016; **?**; Anderson, 2018; Bridge, 2016, 2017). At the time of our study, teachers at Bridge schools were less educated than their public and private counterparts. Three-quarters of teachers in public and private schools had acquired more than a secondary school education compared to just under one-quarter of teachers in Bridge schools. Relative to public school teachers, Bridge teachers were younger, less experienced, and more likely to be novice (first-year) teachers. On average, their total compensation amounted to between one fifth and one third of the average public school teachers total compensation and approximately the same as teachers in other private schools serving this population. They worked longer hours, including Saturdays.

Data from pupil surveys indicate that scholarship recipients who enrolled at a Bridge primary school were eight percentage points more likely to report that their playing fields have hazards, compared to 34 percent of scholarship non-recipients attending other primary schools. Recipeints who enrolled at Bridge were six percentage points less likely to report that their teachers engaged in corporal punishment, compared to 83 percent of scholarship non-recipients. Among scholarship non-recipients (the control group), pupils who chose to attend Bridge were more likely to report missing school due to unpaid school fees than pupils in public schools.

Subsequent to the period analyzed in our study, Bridge's parent company NewGlobe reduced

the number of private schools operated by Bridge from 405 to 112, and launched a new model in which it primarily acts as a service provider to governments. Under this model, which now accounts for the bulk of students reached by NewGlobe, teacher qualification, compensation, and working conditions follow standard public sector guidelines; governments similarly set curricular, school infrastructure, and child safety standards, and costs of standardization are covered by the state rather than through fees to parents.

Our results suggest that policymakers may wish to explore incorporating more structure and standardization into their education systems. Since the effects of standardization in education could vary between public and private schools, as well as with local circumstances and implementation, careful evaluation of efforts to introduce more structure into public education systems would be needed.

The analysis conducted in the paper was outlined in the *Preliminary Analysis and Plan for Subsequent Analysis* (PAPSA) document (Gray-Lobe et al., 2020). The PAPSA examined attrition rates, covariate balance, and the first stage effect of the scholarship offer on attendance. We developed a detailed analysis plan that discussed the framework for analysis, outcomes, and interpretation, taking into consideration preliminary results (e.g., scholarship take-up) while maintaining constraints on additional data access. This process was intended to ensure that decisions regarding data sources, cleaning, and methods for subsequent analysis were made without knowing what influence those decisions might have on estimated treatment effects on outcomes to be analyzed in the next stage.[5]

This paper contributes to several different literatures. Broadly, we contribute to discussions on determinants of learning in low and middle-income countries (Evans and Popova, 2016; Kremer et al., 2013) using a large experimental study. In addition to a large nationwide sample, the study examines student achievement across all eleven pre-primary and primary grades (up to eighth grade). The paper also contributes to the literatures on scripted instruction (Piper et al., 2014, 2018; Stockard et al., 2018; Eble et al., 2021), monitoring (Duflo et al., 2012), non-civil service teachers (Duflo et al., 2015a; Muralidharan and Sundararaman, 2013), and the role of technology in education (Muralidharan et al., 2019; Cristia et al., 2017). We also contribute to the literature

[5]This sequential preliminary analysis process is a variation on procedures discussed by Olken (2015) and implemented by Leaver et al. (2021) and Humphreys et al. (2013).

on non-state (or private) education providers (Hoxby and Murarka, 2009; Abdulkadiroğlu et al., 2011; Cohodes et al., 2019; Angrist et al., 2002; Hsieh and Urquiola, 2006; Dean and Jayachandran, 2019; Muralidharan and Sundararaman, 2015; Bold et al., 2011; Romero et al., 2020; Zuilkowski et al., 2020; Fazzio et al., 2021).

Our study relates to the literature on the effects of preschool for child cognitive development and school preparedness (e.g. Currie, 2001; Baker et al., 2008; Weiland and Yoshikawa, 2013; Lipsey et al., 2018; Gray-Lobe et al., 2021) and a growing literature on the determinants of early childhood education quality in developing countries (Blimpo et al., 2018; Huillery et al., 2017; Dean and Jayachandran, 2019; Ngware et al., 2018). One distinctive feature of our study is that it compares a highly structured approach to early childhood development (ECD) to less structured ECD schools, unlike most studies where results are often driven by large effects among those whose counterfactual option was no schooling.[6]

Topically related to our study, though contextually distinct, are Romero et al. (2020) and its follow-up (Romero and Sandefur, 2021), which present experimental evidence on the effect of contracting management of public schools in Liberia to private providers, including Bridge.[7] In the Liberian context, Bridge had more management authority than under its current operations as a service provider to Ministries of Education and could limit enrollment and choose not to work with certain teachers, as emphasized by Romero et al. (2020), but much less than under the private school model we study, where Bridge controls all aspects of the school, including hiring and firing decisions, teacher incentives and accountability measures, class size, and the amount of classroom instruction. In the model we study, Bridge thus had the ability to make a series of potentially complementary management changes.

Our work also connects to the literatures on scarcity of large private firms in low- and middle income countries and the relative productivity of such firms (Hsieh and Klenow, 2009; Hsieh and Olken, 2014), on standardization of production processes more broadly (Marshall, 1919; Arrow, 1974; Choi et al., 2016; Haynes et al., 2009; Eifert et al., 1997; Wilson, 1996; Autor, 2014) and on

[6] A common empirical challenge in the early childhood education literature is the multiple counterfactual issue. This arises because parents often have many childcare options of varying quality, including homecare (Berkes and Bouguen, 2022; Dean and Jayachandran, 2019; Kline and Walters, 2016).

[7] The education systems in Liberia and Kenya are very different: Kenya's education system is one of the strongest in Sub-Saharan African whereas Liberia's is relatively weak and the country is recovering from a recent civil war (Sandefur, 2016; Uwezo, 2016; Martin and Pimhidzai, 2013).

private-sector (personnel) management practices (Bloom et al., 2015; Lemos et al., 2021).

The rest of this paper is structured as follows: Section 2 provides background information on the Kenyan education system and describes Bridge's approach to standardizing education and its strategy for doing so through private schools at the time of the study. Section 3 discusses the data collection, the scholarship program, sampling procedures, and assessment design. Section 4 presents the empirical framework to guide the interpretation of results. Section 5 discusses the impact of scholarship receipt and Bridge enrollment among the main sample of pupils who were judged most likely to attend public schools in the absence of a Bridge scholarship including effects on subject-matter tests, non-subject matter outcomes, and timely grade progression. Section 6 discusses the effect of Bridge enrollment on intermediate inputs and classroom experience. Section 7 argues that test score impacts were not driven by income effects of scholarships and Section 8 examines the effects of Bridge relative to other private schools using a smaller sample of pupils judged likely to have attended other private schools if they did not win a scholarship to Bridge. Section 9 discusses concerns raised by civil society and labor organizations regarding Bridge's private school model, subsequent policy steps taken by various stakeholders, and NewGlobe's move from a private school business model to one based on selling services to governments. Section 10 summarizes and concludes by discussing broader implications for standardization in education.

2 Background

This section provides background context on the Kenyan education system (Subsection 2.1); describes the standardization strategy followed by Bridge (Subsection 2.2); implications of Bridge's private-school model (Subsection 2.3), including teacher qualifications and compensation (Subsection 2.4).

2.1 The Kenyan Education System

This section provides a brief background on the Kenyan education system. As in many other lower-middle-income countries, access to primary education has expanded rapidly in Kenya, with public education statutorily free at the primary school level.[8] Almost all children attend primary

[8]Parents of public primary school pupils may still pay ancillary fees for uniforms or for additional teachers hired on short-term contracts (Evans and Ngatia, 2020; Zuilkowski et al., 2017).

school. At the end of primary school, pupils take a high stakes primary school leaving exam, the Kenya Certificate of Primary Education (KCPE) exam, which tests knowledge in math, Kiswahili, English, science, social studies, and religious education. Performance on this exam determines which secondary school, if any, admits the student.

At the time of the study, primary school education consisted of eight grades, known as standards. Pupils who completed Standard 8 could sit for the KCPE. The Kenyan school year runs from January to November and each academic year is divided into three terms. Pre-primary schooling in Kenya consisted of three grades - Baby class, Nursery class, and Pre-unit - intended to serve three-, four- and five-year old pupils, respectively.

Public school teachers are employed by the national government, are subject to civil service protection, and earn much more relative to GNP per capita than those in higher income countries. This is likely in part because they are much more educated than the typical worker, typically having specialized post-secondary education, and because they have historically had a strong union, with salaries for public-school teachers several times as large as those for teachers in private schools. The majority of teachers are employed through the Teachers Service Commission (TSC). The minimum requirement for public primary school teachers is a certificate in primary education, which is a two-year (plus three-week practicum) post-secondary credential.

Private school education has expanded rapidly in recent decades. Approximately 33 percent of pre-primary school pupils and 16 percent of primary pupils in Kenya were enrolled in private schools (Ministry of Education, 2016a).[9] Aside from a few elite primary schools, most are "mom and pop" operations, with a single establishment, rather than chains.

Historically, many pre-schools were organized informally at the local community level and therefore did not fit neatly into the categories of "public" and "private". In particular, while many pre-primary schools met on the premises of public primary schools, teachers were not centrally hired civil servants paid by the national government, and instead supported themselves by charging parents modest fees. With the adoption of the new Kenyan constitution in 2010, pre-school was declared a responsibility of the new county governments. However, there is still substantial heterogeneity in the extent to which already existing pre-schools have been formalized (Devercelli

[9]The increase may have been driven partly by the Free Primary Education Program of 2003 that led to higher pupil-teacher ratios and fewer resources available per pupil, which, in turn, may have led some families to seek out private schools. (Lucas and Mbiti, 2012; Bold et al., 2014)

and Sayre, 2016).

Teachers have low levels of content and pedagogical knowledge and high rates of teacher absence (Bold et al., 2017; Martin and Pimhidzai, 2013). Only 40 percent of primary school teachers in Kenya could demonstrate minimum content knowledge required to teach primary school, and only 39 percent could grade a fourth grader's test with greater than 80 percent accuracy (though Kenyan teachers were the best among several Sub-Saharan African countries studied along this dimension viz. Mozambique, Nigeria, Senegal, Tanzania, Togo, and Uganda). Kenyan teachers were absent from class during 47 percent of unannounced visits and spent about two hours and twenty minutes a day teaching (Bold et al., 2017; Martin and Pimhidzai, 2013).

Although higher than most other countries in Sub-Saharan Africa, learning outcomes in Kenya are substantially lower than those in high-income countries (World Bank, 2020). In third grade, only 47 percent of Kenyan pupils could solve a mathematics problem at the second grade level and 40 percent could read English at the second grade level (Uwezo, 2016).

Since the time of the scholarship program the Kenyan public education system has changed substantially, with the introduction of a new Competency Based Curriculum (CBC) in primary schools. This was, in part, due to concerns that the previous curriculum encouraged rote learning. The CBC instead emphasizes a broader set of skills, as well as learner-centered instruction and parental engagement.

2.2 Standardization of education in Bridge

While some degree of structure and standardization is often present in education – for example, in the use of standardized curricula and textbooks – standardization at Bridge is both deeper and broader than is typical in education. For example, as discussed below, Bridge attempts to more fully standardize instruction by articulating (in detailed "lesson guides") precisely what teachers should teach and how they should teach it, by providing ready access to these "guides" through tablet computers (Subsection 2.2.1), and by promoting consistent use of the guides through frequent monitoring and feedback (Subsection 2.2.2). Bridge similarly standardizes other elements of its systems such as physical construction of schools and day-to-day financial transactions (Subsection 2.2.3). Much of this is facilitated through the use of technology. Like many other firms, from Ford Motors to Uber, which seek to standardize production, Bridge uses technology to facilitate both

activities (Marshall, 1919; Autor, 2014). Unlike some education technology firms, Bridge provides technology to teachers and school heads, not directly to the pupils.

2.2.1 Lesson guides

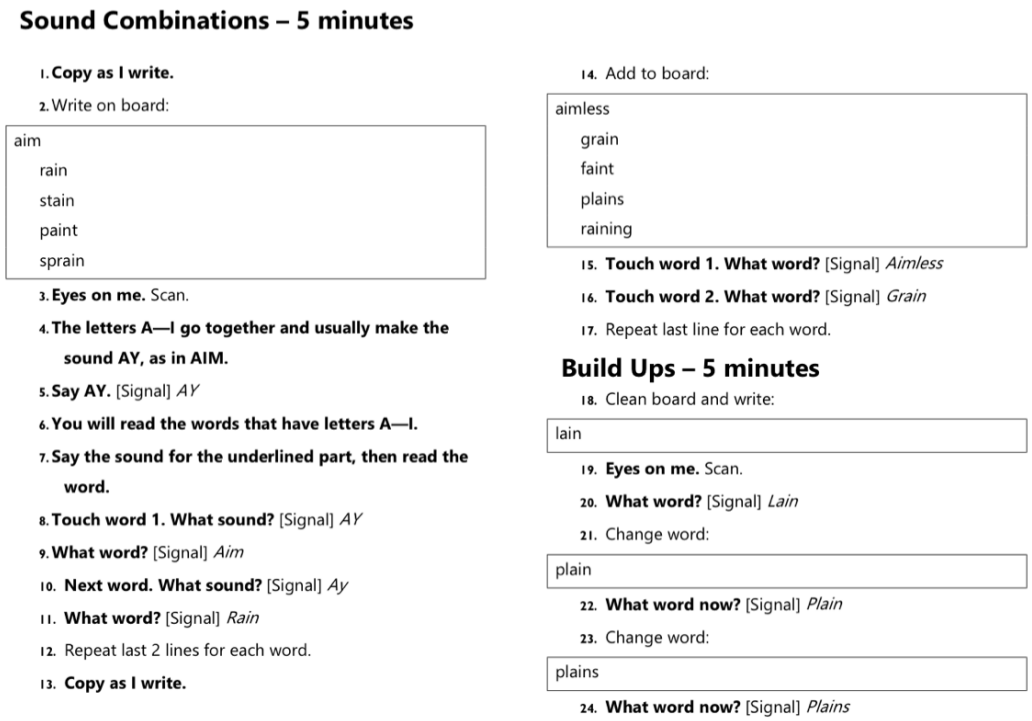
Bridge centrally designs highly-detailed lesson guides in an effort to fully articulate teachers' role in classroom instruction. These lessons are shared with teachers via tablet computers.

These lesson guides contain detailed, step-by-step instructions for what teachers should say along with cues for teacher actions such as writing on the board, calling on pupils, and providing feedback and encouragement. Instructions are highly granular, including for example cues telling teachers when to erase the blackboard, when to walk around the class, when to praise pupils, and when to tell pupils to close their books after the lesson (Riep and Machacek, 2016).

Figure 1 reproduces a portion of a typical lesson guide used in Bridge schools. Things that teachers are supposed to say are in bold; instructions to the teachers referring to their training are in plain text. For example, when the instruction "scan" appears, it means that teachers should "notice what each pupil is doing by looking at them and moving around all parts of the room." The phrase "signal" tells the teachers they should use a particular gesture taught during training. The pupils are taught that when the teacher says "eyes on me," they are supposed to look at the teacher (See Appendix A for the full lesson guide).

Bridge teachers follow a set of instructions for dealing with behavioral management in the classroom. Figure 2, shows examples of techniques taught to teachers. Examples of instructions given to teachers include "check, respond, leave," which involves checking one pupil's work, providing quick verbal feedback, and moving on to check another pupil's work and "leading a cheer," which involves a teacher encouraging the class to do one of a set of cheers, typically combining verbal and physical signals to acknowledge a pupil who answered a question correctly or behaved well. Note that these classroom management instructions may be particularly important for the early career teachers common in Bridge classrooms, since there is some evidence early career teachers often lack strong classroom management skills (Ingersoll and Smith, 2003; Johnson et al., 2012).

Figure 1: Excerpt from a typical lesson guide used in Bridge schools



Sound Combinations – 5 minutes

1. **Copy as I write.**
2. Write on board:

aim
rain
stain
paint
sprain

3. **Eyes on me.** Scan.
4. **The letters A—I go together and usually make the sound AY, as in AIM.**
5. **Say AY.** [Signal] *AY*
6. **You will read the words that have letters A—I.**
7. **Say the sound for the underlined part, then read the word.**
8. **Touch word 1. What sound?** [Signal] *AY*
9. **What word?** [Signal] *Aim*
10. **Next word. What sound?** [Signal] *Ay*
11. **What word?** [Signal] *Rain*
12. Repeat last 2 lines for each word.
13. **Copy as I write.**
14. Add to board:

aimless
grain
faint
plains
raining

15. **Touch word 1. What word?** [Signal] *Aimless*
16. **Touch word 2. What word?** [Signal] *Grain*
17. Repeat last line for each word.

Build Ups – 5 minutes

18. Clean board and write:

lain

19. **Eyes on me.** Scan.
20. **What word?** [Signal] *Lain*
21. Change word:

plain

22. **What word now?** [Signal] *Plain*
23. Change word:

plains

24. **What word now?** [Signal] *Plains*

2.2.2 Monitoring and feedback

Another key element of standardization, besides articulating and defining production processes, is the development of systems for monitoring workers and providing feedback. These activities are complementary: clearly articulating production processes creates a framework against which to monitor worker activity; monitoring and feedback help ensure production processes are followed as intended or fixed when they fall short.

At Bridge, school heads are given detailed rules and their own scripts to use in checking teachers' compliance with lesson plans and other aspects of Bridge procedure. School heads are trained to observe teachers to quickly check on teachers twice daily, recording whether the teacher is actively leading the classroom, whether he is on the correct lesson, and whether all the pupils have the relevant material. If this quick check reveals problems, the school heads follow instructions about helping the teacher. In addition, the school head conducts one full-length classroom observation daily. For the classroom observation, school heads are instructed to follow six steps in these observations: prepare, observe, give feedback, follow up, give follow-up feedback, and enter scores

Figure 2: Standardization of classroom management: Examples of techniques taught to teachers

- **Passing Procedure:** This is the process used to pass all learning materials to the pupils during the lesson. The teacher gives all of the row's materials to the passer. The passer is the pupil on the right side (the teacher's left when facing the pupils) of the first desk in each row. The passer takes the correct amount of materials for her row and passes the rest over her head to the person directly behind. The same process is repeated until the last bench gets the learning materials.

- **STRIVE:** This is an acronym with the letters meaning: S – Sit Tall, T – Track the speaker with your eyes and body, R – Respond to Questions, I – Inquire, V – Visualize Success and E – Ears all listening. The teacher says these letters to reset the class by asking the pupils what one of the letters mean. For example, a teacher could say, "What does "T" stand for?" and all the students will say, "Track the teacher with your eyes and body." Then the students will track the teacher with their eyes and body. **STRIVE** is used in Lower Primary and Upper Primary (G1-C8)

- **STEP:** This is an acronym with the letters meaning: S – Sit Tall, T – Track the speaker with your eyes and body, E – Ears all listening and P - Participate. The teacher says these letters to reset the class by asking the pupils what one of the letters mean. For example, a teacher could say, "What does "T" stand for?" and all the students will say, "Track the teacher with your eyes and body." Then the students will track the teacher with their eyes and body. **STEP** is used in ECD classes: Baby Class, Pre-primary 1 and Pre-primary 2.

- **Scan**: Scan means noticing what each pupil is doing by looking around all parts of the classroom. When there is enough time, teachers can move around the room to make scanning more effective. Scanning helps teachers catch pupil misbehavior and reminds pupils that the teacher is paying attention to their actions. This strategy is used to motivate pupils to behave and work hard.

- **Narrate the positive**: Narrating means naming a student and stating their good behavior or academic effort. Teachers should make sure that they are calling on pupils who are behaving and setting a good example for others. Teachers should narrate the positive anytime they want all pupils to do something, but a few are not doing it. Teachers might narrate the positive after giving an instruction or when pupils are working independently. This helps teachers create a positive classroom environment where pupils feel comfortable to participate.

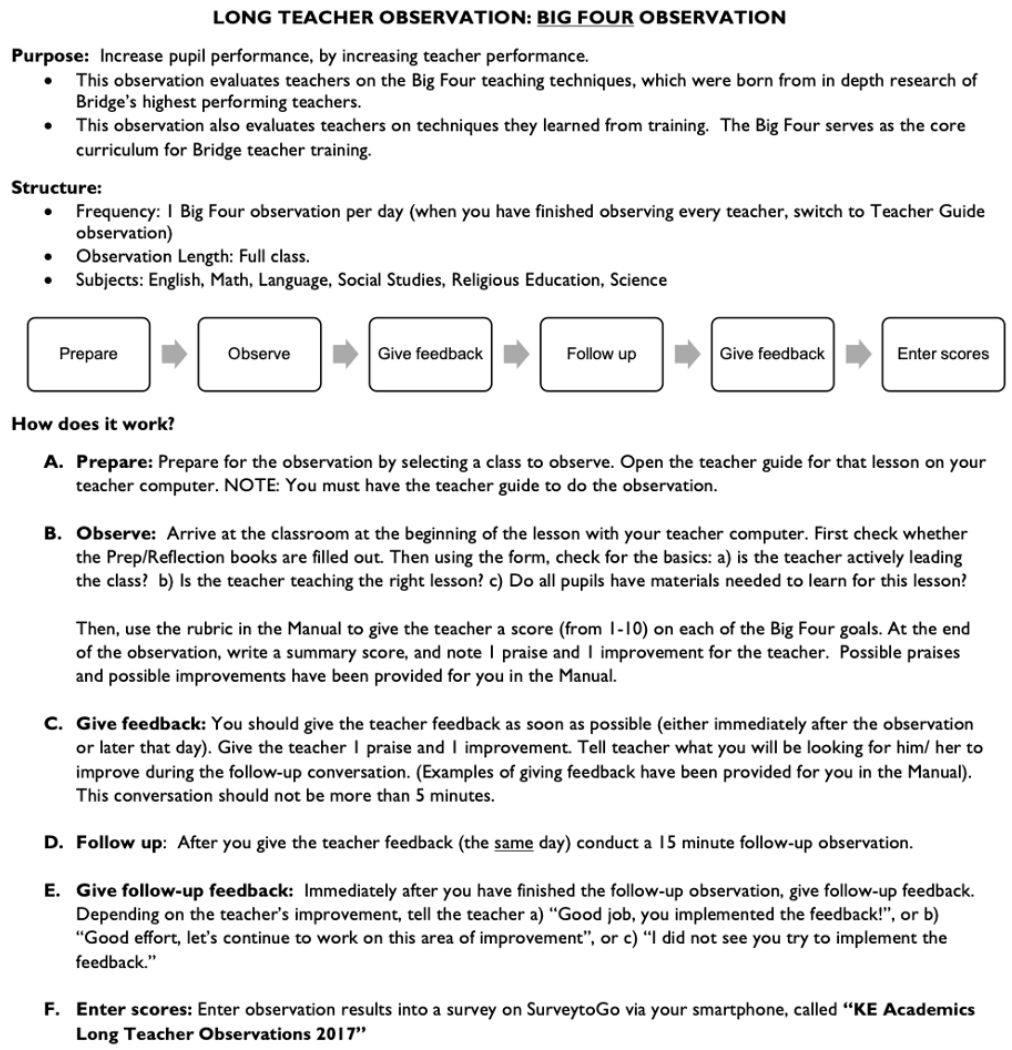
- **Cold Call:** Cold calling is when a teacher calls on a pupil to answer a question regardless of whether they are volunteering or not. It is a technique that maximizes student engagement. Its premise is simple: teachers use cold call to call on any student, at any time. A teacher needs to ask the question, then name a pupil, then signal. Teachers strategically switch between cold calling on different types of pupils: pupils who are struggling, pupils who do not volunteer often because they are shy and pupils who are average and high performers.

- **Check Respond Leave: Check** means to determine if a pupil's work is correct or incorrect, **Respond** means to give the pupil clear verbal feedback. If a pupil is correct, say "correct" and quickly put a tick mark. If a pupil is incorrect, say "incorrect" and tell them what errors were made and **Leave** means to walk away quickly to check another pupil's work.

- **Redirect:** Redirect means naming a pupil and saying how you want their behavior or academic effort to change. When we redirect a pupil, we name the pupil and say precisely how you want their behavior or effort to change. Using a redirect can help pupils stay on task. Redirect should follow narrating the positives. Example: Paul is tapping his feet under his desk. The teacher would say, **"Paul, feet still."**

- **Cheers:** Cheers are celebratory actions. The teacher may request the pupils to say a cheer to celebrate a pupil or pupils who have done great work or behaved well. Cheers include:
    - Silent Cheer, Parapanda Cheer, Roller Coaster Cheer, Banana Cheer, Mavellous Cheer, Shake your Collar Cheer, Coca Cola Cheer etc.

Figure 3: Standardization of monitoring: Excerpt from a teacher observation guide



LONG TEACHER OBSERVATION: BIG FOUR OBSERVATION

Purpose: Increase pupil performance, by increasing teacher performance.

- This observation evaluates teachers on the Big Four teaching techniques, which were born from in depth research of Bridge's highest performing teachers.
- This observation also evaluates teachers on techniques they learned from training. The Big Four serves as the core curriculum for Bridge teacher training.

Structure:

- Frequency: 1 Big Four observation per day (when you have finished observing every teacher, switch to Teacher Guide observation)
- Observation Length: Full class.
- Subjects: English, Math, Language, Social Studies, Religious Education, Science

Prepare → Observe → Give feedback → Follow up → Give feedback → Enter scores

How does it work?

A. Prepare: Prepare for the observation by selecting a class to observe. Open the teacher guide for that lesson on your teacher computer. NOTE: You must have the teacher guide to do the observation.

B. Observe: Arrive at the classroom at the beginning of the lesson with your teacher computer. First check whether the Prep/Reflection books are filled out. Then using the form, check for the basics: a) is the teacher actively leading the class? b) Is the teacher teaching the right lesson? c) Do all pupils have materials needed to learn for this lesson?

Then, use the rubric in the Manual to give the teacher a score (from 1-10) on each of the Big Four goals. At the end of the observation, write a summary score, and note 1 praise and 1 improvement for the teacher. Possible praises and possible improvements have been provided for you in the Manual.

C. Give feedback: You should give the teacher feedback as soon as possible (either immediately after the observation or later that day). Give the teacher 1 praise and 1 improvement. Tell teacher what you will be looking for him/ her to improve during the follow-up conversation. (Examples of giving feedback have been provided for you in the Manual). This conversation should not be more than 5 minutes.

D. Follow up: After you give the teacher feedback (the same day) conduct a 15 minute follow-up observation.

E. Give follow-up feedback: Immediately after you have finished the follow-up observation, give follow-up feedback. Depending on the teacher's improvement, tell the teacher a) "Good job, you implemented the feedback!", or b) "Good effort, let's continue to work on this area of improvement", or c) "I did not see you try to implement the feedback."

F. Enter scores: Enter observation results into a survey on SurveytoGo via your smartphone, called **"KE Academics Long Teacher Observations 2017"**

(see Figure 3). They focus on four themes: how well teachers motivate pupils, follow the teacher guide, check on pupils' performance, and respond to pupils' performance.

Teachers' tablet computers also play a role in monitoring. At the start and end of lessons, Bridge teachers use their tablets to check in and out, which helps Bridge track teacher attendance and monitor lesson completion rates.

Bridge also frequently monitors student learning to gather additional data to monitor the performance of teachers and evaluate the effectiveness of their teaching materials. Bridge conducts standardized assessments of student learning up to seven times a year in each subject to track student progress, inform in-service teacher training, and refine lesson guides.

Figure 4: Standardization of monitoring: Teacher guide observation rubric

Teacher Guide Observation Rubric

Note: This is a summary; please see the Manual for more details.

Add	• Definition: Teacher says something that is not written in the Guide. • Tally if: Teacher says something that is not written in the Guide related to content or facts, or extra comments that last longer than 10 seconds.
Repeat	• Definition: Teacher repeats a line multiple times. • Tally if: Teacher says a line 3 times or more.
Skip	• Definition: Teacher skips a line or action in the Teacher Guide that he/she is supposed to say or do • Tally if: Teacher skips a line he/she is supposed to say, or skips any directions from the Teacher Guide about the blackboard, materials, teacher actions and pupil actions, etc.
Rephrase	• Definition: The teacher changes or rephrases lines in the Teacher Guide, including any type of translation. • Tally if: The teacher changes or rephrases lines in a significant way that changes facts content, or activities, OR translates the Guide.
Excessive Reset	• Definition: Teacher does STEP/STRIVE or resets the class too many times. • Tally if: Teacher does STEP/STRIVE or resets the class more than 4 times during class. (After a teacher's first 4 Whole Class Resets, tally an error for each additional Whole Class Reset the teacher uses during the lesson you are observing. The first error should be tallied on the teacher's 5th reset.)

To calculate the Teacher Guide accuracy percent:

$$\frac{\text{(End line-Start line)-Number of errors}}{\text{End line-Start line}} \times 100$$

Depending on the score, feedback should vary.

Green: Teacher scores 90-100%. Tell the teacher one praise.

Yellow: Teacher scores 80% to 89%. Tell the teacher one praise and one area of improvement (based on the most common type of inaccuracy).

Red: Teacher scores 79% and below. Tell the teacher one area of improvement (based on the most common type of inaccuracy).

Teacher Guide Observations **Teacher Name:** _____ **Class:** _____

		Line Started	Addition	Repeat	Skip	Rephrase	Reset	Calculate Score	Feedback	15-min Follow- up Observation
Date:								*Total Number of Lines*	*One Praise*	*Did teacher apply feedback? (Circle one)*
Subject:										
Actively leading?	YES / NO	**Line Ended**						*Number of Errors*	*One Improvement*	Yes
Correct lesson?	YES / NO									
Materials available?	YES / NO							*Final Score (%)*		Somewhat
Filled out Prep & Reflection Book?	YES / NO									No

2.2.3 Standardization of school management and infrastructure

Procedures for financial decisions are centralized and hence standardized, allowing for monitoring and limiting discretion of academy managers. Bridge uses electronic payment systems such as M-Pesa for all financial transactions including fee payments by parents, staff salaries, and payments to other school suppliers. Standardized procedures delineate how long school fee payments can be delayed before pupils are not allowed in class.[10] School managers use their Bridge-issued smartphones to process staff and supplier payments without handling any money themselves. Standardization of financial transactions helps avoid the logistical issues of dealing with cash payments, allows real-time access to financial data, and reduces the scope for financial irregularities. This may free up the school head's time for monitoring and providing feedback to teachers.

Bridge also standardizes physical infrastructure, allowing it to buy supplies in bulk, limit agency problems created by delegating decisions to local construction managers, and hold down costs. Bridge classrooms use a wooden framing and are enclosed by iron sheeting. Public schools, on the other hand, are required to have stone, brick, or concrete walls. Bridge's founders recognize that the model deprioritizes physical infrastructure and they have argued that this frees up resources for expenditure on other inputs that can improve school quality (Beck and May, 2016). According to their CEO, the schools are designed so they can be built with "the kind of instructions that you would see at Ikea. Because of this, it is easy to assemble a school and to add to a school" (Rangan and Lee, 2010). Standardization extends to the construction of latrines, blackboards, the arrangement of the classroom, the color of the academies, and land procurement (Rangan and Lee, 2010).

The standardized production of classroom facilities also has features that may be complementary to the monitoring of teachers. The side of the classroom with the door features an open window that spans the length of the classroom so that an observer standing outside the classroom can observe whether a teacher is in a classroom and teaching. The fact that activity within classrooms can easily be heard in other classrooms may also contribute to monitoring. Note that this feature comes with a tradeoff, as there is usually another classroom on the other side of this open window, which may make the classroom environment noisy.

[10] There are sometimes exceptions in practice.

Bridge's efforts to more fully standardize education – through lesson guides combined with intensive teacher monitoring and feedback – likely did not lead to complete standardization in practice. According to a survey conducted by Education International and KNUT (2016), 52.4 percent of Bridge teachers reported that they always follow the lesson guides, 14.3 percent said they follow the instructions most of the time, and 33 percent said they follow the lesson guides "from time to time."

2.3 Implications of standardization for scale

Business models based on developing standardized procedures involve fixed capital costs of developing, testing, and refining their standardized procedures. Organizations attempting to standardize production must therefore achieve sufficient scale to amortize these fixed costs over a large number of users. Conversely, standardization facilitates operation at scale over multiple locations, and thus can be seen as a technique for expanding the span of control of a limited number of highly-skilled workers charged with developing and refining the standardized procedures. Standardization and scale are thus strong complements.

In principle, scale could be achieved either through government provision or by operating many private schools. At the time of this study, Bridge sought to achieve scale by operating hundreds of private schools with hundreds of thousands of pupils. Doing so in a lower-middle income country in turn implies targeting non-elite households, and thus limiting costs that scale with enrollment.

Like McDonald's, Walmart, and many other firms using standardization, Bridge is targeted to a mass market, not to consumers at the high end of the income distribution. Bridge charges annual fees around KES 9,000 (less than US$ 100) for primary school, or about six percent of GDP per capita (Table 1). Bridge schools are often located in informal settlements in major cities and other schools are located outside of major cities altogether. The populations served by Bridge tend to be non-elite.

The decision to scale with a private school model in a lower-middle income country meant that, to be financially viable, Bridge needed to hold down recurrent costs that scaled with the number of pupils. While in principle standardization can be used solely to improve quality without cutting labor costs, standardization can also be seen as complementary with hiring cheaper workers with less education and experience because a limited number of very high-skill workers develop the

standardized procedures, and thus the job of teachers becomes simpler, since they simply need to follow a lesson plan, rather than devise their own lesson plans.

2.4 Teacher recruitment, qualifications, and pay

The largest share of expense in most education systems is teacher salaries, so holding down variable costs has implications for teacher HR practices. Standardization can enable firms to hire workers with less formal education or sector work experience to perform production tasks. This may make it possible to pay workers less, require longer hours, and apply more intensive monitoring than under artisanal production (Knutsson and Tyrefors, 2022).

According to Bridge's administrative data, in 2017 only 23 percent of Bridge's primary school teachers had a teaching certificate at the time of hire, and the remainder had no post-secondary qualification (Table 1).[11] By not requiring post-secondary credentials, which typically represent a smaller share of the labor force in lower-middle income countries, Bridge has been able to draw from a larger pool of secondary school graduates. The teacher recruitment process uses centrally-developed testing and interview procedures to screen candidates.[12]

New teachers at Bridge receive a ten-day training to introduce them to the tablet computers and provide instruction and practice on delivering lessons from the tablets as well as on entering student performance data. Teachers are also trained on how to engage and interact with pupils, check their work, and provide feedback to pupils and their parents. During the training, learning is gauged through quizzes, the delivery of mock lessons, and a final written assessment. The training covers Bridge policies such as following national guidelines prohibiting corporal punishment. The training also includes a module on marketing and recruiting new pupils (Education International and KNUT, 2016).

At the time of the study, staff at Bridge schools – which include teachers and one academy manager – received far lower salaries and benefits than those in public primary schools. Bridge teachers were paid around KES 10,100 per month. Bridge Academy managers were paid around

[11] This figure has been rising since a 2016 change in government regulation (Ministry of Education, 2016b). In 2016, the first year of the scholarship program, 23 percent of new hires had teaching certificates at the time of hire. By 2019, 57 percent of new hires had teaching certificates.

[12] Although newly-hired teachers are often recruited locally, the Bridge hiring process is centralized and involves testing candidates in subject knowledge, pedagogical knowledge, and other topics. Bridge invites select candidates to a training where they are further evaluated and only a subset of the candidates is finally hired.

11,100. All staff received housing allowances equivalent to 15 percent of their salary.[13] Meanwhile, Public primary teachers (excluding headteachers) were paid between KES 30,000 and 50,000 in direct compensation (base wages, housing allowance, commuter allowance) (TSC-KNUT CBA, 2016). Headteachers and deputy headteachers in public primary schools receive between KES 40,000 and 86,000 per month in salary (TSC-KNUT CBA, 2016). Additionally, public school teachers can look forward to pension and health benefits. We estimate that the average public expenditure per public primary school teacher was KES 56,300 per teacher per month in total compensation using figures from the Kenyan Treasury (The National Treasury, 2018), or around five times what a Bridge teacher earns, including their housing allowance.[14]

Bridge teachers not only receive lower wages, but also work longer hours. Using pupil reports from the endline survey, we find that Bridge features a longer school day and more Saturday school than public options at both the primary and pre-primary levels (Table 1).

Less experienced and educated teachers could be hired at lower wages, reducing the cost of operating with lower class size. The average PTR in primary classrooms in Bridge schools is 20 compared to 34 in public primary schools and 17 in other private schools. At the pre-primary level, Bridge has smaller classes than both public and other private schools. The PTR is 13 in Bridge schools, 31 in public pre-primary schools and 25 in private schools (Table 1).

Data shared with the research team by Bridge showed that in the year prior to our study, many teachers were replaced. Many of these separations from Bridge may have been the employees' choices to leave their jobs, but a sizable share of these separations stemmed from Bridge's decision as an employer, whether for reasons of redundancy, absenteeism, or other violations of Bridge's professional guidelines for its staff.

Combining information on PTR and compensation indicates that for primary grades, teacher

[13]Bridge staff also receive bonuses that are linked to enrollment in their schools. Teachers can receive enrollment bonuses up to 25 percent of their base salary, and academy managers can receive bonuses up to 165 percent from their base level based on the number of pupils recruited (Education International and KNUT, 2016). We do not have data on the average amount paid for these bonuses. Because the bonuses were paid contingent on high levels of enrollment that were rare in Bridge schools at the time of the study, we believe these bonuses were unlikely to account for a large portion of average compensation.

[14]This is calculated by dividing the amount paid for Primary school level "Teacher Resource Management – compensation" (KES 1.453 billion) by the number of active primary school teachers in 2017 (145,317) (The National Treasury, 2018).

costs per pupil in Bridge schools are a third of that in public schools (Table 1).[15]

While standardization allows for the hiring of workers with less experience and formal qualifications, it also creates the need to hire a (relatively) small number of highly skilled workers whose job is to standardize the production processes. Bridge's payroll also includes some high-skilled workers, sometimes based in the United States, who develop the standardized practices. Their work involves the application of cognitive competencies, rather than standardized processes, and are paid many multiples of what a teacher or academy manager earns at Bridge.

3 Data, Sample, and Study Design

This section discusses the data collection process (Subsection 3.1), the scholarship program including the application and recruitment process (Subsection 3.2), the classification of applicants based on where they would likely enroll if they did not win the scholarship (Subsection 3.3), the final samples and associated descriptive statistics (Subsection 3.4), and the design of subject knowledge tests used for assessments and the construction of outcomes (Subsection 3.5).

3.1 Data

Pupil and parent level data analyzed in this study come from four main sources: (i) the scholarship application files of pupils in late 2015; (ii) phone call surveys with parents in 2016, 2017, and 2019; (iii) an in-person endline survey at the end of 2017. The scholarship application files provide information on the background and demographic characteristics of pupils and their caregivers and their previous contact information. Application data also include the percent correct from the pupil's most recent end-of-term test.[16] The phone call surveys that followed in 2016 and 2017 were undertaken to maintain up-to-date contact information to facilitate the endline survey and collect information on the type of school the pupil was attending.

[15]The ratio of teacher costs per pupil between Bridge schools and public schools is given by:

$$\frac{\text{Compensation in Bridge/PTR in Bridge}}{\text{Compensation in public schools/PTR in public schools}}$$

The average teacher compensation for primary grade teachers is US$ 116 in Bridge schools and US$ 563 in public schools (Table 1). (For Bridge teacher compensation, we include both their monthly base salary and a 15 percent housing allowance.) According to this approach Bridge's teacher compensation per pupil is 35 percent that in public primary schools.

[16]These tests were developed by the pupil's teacher and are not comparable for pupils across schools.

The endline survey was conducted in households between November 2017 and March 2018. It consisted of interviews with pupils and parents.[17] The parent module collected data on household education expenditures, missed classes, parental engagement in pupil education, satisfaction with the school, and other outcomes. The pupil module included a subject knowledge assessment with sections on math, Kiswahili, English, science, and social studies. The pupil survey also included tasks measuring general cognitive and non-cognitive skills and items related to the pupil's experience in school, including teacher behavior and school facilities. This was followed by a phone call survey in 2019 that collected data on grade progression, secondary school transitioning, and the primary school exit exam (KCPE) outcomes. Appendix Table A1 presents the timeline of the data collection activities and key information collected in each survey. Additional details on considerations that influenced specific data collection choices can be found in Gray-Lobe et al. (2020).

3.2 The scholarship program

The NGO UnitedWeReach (UWR) provided scholarships for the program.[18] Children intending to enter any grade in pre-primary and primary school were eligible to apply; however the study includes only applicants up to grade 7 (pupils who would start grade 8 in 2016 were ineligible). Applicants indicated up to two academies that they would enroll at if they received the scholarship. Applications were received from all 405 Bridge academies then operating in Kenya.

Initially, scholarships were provided for one academic year, starting in January 2016. In August 2016, after additional funding was secured, scholarship recipients were informed scholarships would be renewed for the 2017 school year.[19]

3.2.1 Scholarship application recruitment and procedures

The scholarship program was advertised locally in November 2015: Bridge publicized the scholarship program to parents who were identified through its normal recruitment process; Bridge also encouraged parents of current Bridge pupils to apply and to refer others to apply. Applications were accepted through an internet portal or through a dedicated call-center. In a parallel process,

[17] The parent survey was conducted with an available caregiver, not necessarily a biological parent. However, because in most cases, the respondent was a biological parent, we refer to these respondents simply as "parents".

[18] One of the Bridge co-founders was a board member of UWR in 2015.

[19] In January 2018, scholarships were extended for an additional one term for a subset of 2,266 pupils. We find that this extension had only a small effect on Bridge enrollment, and so we ignore it in this analysis.

staff from Innovations for Poverty Action (IPA) conducted door-to-door outreach in 100 randomly chosen Bridge school locations to provide information on the scholarship program.[20] A total of 29,949 applications were received.

Some applications were removed prior to randomization because they did not satisfy eligibility criteria established by the scholarship donor or because applicants had received scholarships through other contemporaneous scholarship programs administered by Bridge.[21] 3,534 applications were removed because they did not meet eligibility criteria.[22] 515 applications were removed because the pupils had received scholarships through merit-based programs administered by Bridge. After applying these restrictions the total number of applications was 25,898.

We note that the identities of applicants recruited through local Bridge schools were not verified centrally by Bridge, UWR, or the research team. The scholarships were non-transferable, but the possibility of collusion with local agents (academy managers and teachers) and identity fraud presents a potential threat to the internal validity of the results. During endline activities, no cases of attempted fraud were identified, so we believe that it is rare if it occurred at all. In the Online Appendix, we show that the results reported here are largely unchanged restricting the sample to those who applied through face-to-face meetings with IPA field officers.

3.3 Categorization by probable outside option and randomization

Prior to the lottery, 25,898 scholarship applicants were classified by their probable outside schooling choice and grade level 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.

[20] IPA did not advocate for or encourage people to enroll at Bridge; however, if parents and pupils decided they would like to apply to the scholarship program, IPA provided them with assistance.

[21] See Gray-Lobe et al. (2020) for additional details.

[22] A "selection score" was used to determine eligibility for current Bridge pupils. This score combined information about the pupil's household socioeconomic status, academic achievement, and whether they had siblings at Bridge. 2,590 applications of current Bridge pupils were removed because their index score was below the threshold. All applications needed to demonstrate satisfactory academic performance. To be eligible for the scholarship pupils needed to have scored above 40 percent on their most recent classroom test score (the end of term test from Term 3 of 2015). Only 944 applications were removed due to this requirement, in part because few pupils reported scores below 40 percent and in part due to a data management error. (The application allowed applicants to report their test scores in different fields depending on whether they were reporting an aggregate over multiple subjects or separate subject-specific percentages. The procedure used to determine whether applicants were eligible according to this criterion could have confirmed eligibility via all non-empty test score fields, but instead only used the aggregate test score field, thereby not removing as many applications as it could have).

At the primary school level, pupils were categorized based on whether they were already enrolled at a Bridge school or, if not, whether they planned to enroll at a public or another non-Bridge private school in 2016.

- Sub-group P_{main} consists of pupils who were considered likely to enroll at a public primary school if they did not win a scholarship: those who were not enrolled at a Bridge school at the time of application and who either attended a public primary school at baseline (2015) or stated that they were planning to enroll at a public primary school in the first year of the scholarship program (2016).

- Sub-group P_{brig} consists of pupils who were enrolled at a Bridge school at the time of application or who were matriculating from the Bridge pre-primary program and stated they planned to remain at Bridge even without winning a scholarship.

- Sub-group P_{priv} consists of pupils considered likely to enroll at a non-Bridge private school if they did not win a scholarship; those who were not enrolled at a Bridge or a public primary school at the time of application and who were not planning to enroll at a Bridge or public primary school in the first year of the scholarship program.

At the pre-primary school level, pupils were classified based on whether they were already enrolled at a Bridge school.

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

- Sub-group PP_{brig} consists of pupils who were currently enrolled at a Bridge school at the time of application.

The PP_{main} sample therefore combines pupils who were enrolled or planning to enroll in non-Bridge private programs or public programs. This choice was made in part because at the pre-primary level both public and private school providers charged fees and teachers were often not civil servants (and sometimes simply supported themselves through fee payments) (Nation, 2020). Thus, the distinction between these providers is not as clear as at the primary school level.

The sample was further stratified based on the Bridge school to which the pupil applied, and age group (pre-primary, lower primary, and upper primary); randomization was then conducted

within each stratum. The number of scholarships available in each Bridge school varied across strata, so the probability of a pupil receiving a scholarship also varied across the strata. Where needed, controls to account for differences in probability of scholarship assignment across strata are included in the analysis.

3.4 Sample, attrition, and descriptive statistics

Figure 5 illustrates how we arrived at the final samples. The singleton strata cells were dropped first. We then dropped 343 randomly selected control applications in the PP_{main} and P_{main} samples from one Bridge school as it received a high volume of applications.[23] A subset of pupils in the PP_{brig} and P_{brig} samples were also randomly selected for the endline. Sampling of these pupils was conducted at the level of scholarship randomization strata instead of the individual level, so that entire randomization strata were sampled.

Lamu and Garissa counties could not be visited at endline due to security concerns. However, these applicants are included in phone call follow-up surveys.

During the scholarship program, concerns were raised about the quality of data for applications received through UWR in Meru and Nakuru counties. Several academies in Meru county received unusually high numbers of direct (non-IPA) applications through UWR. In Nakuru, applications submitted through UWR exhibited a very high level of differential attrition. Out of caution, UWR applications in both counties are dropped, leaving only applications submitted through IPA in those counties. This choice was motivated by a concern that anomalies may reflect attempts to defraud the scholarship program, potentially with the intention of selling scholarships for profit. In the Online Appendix, we show that the results are robust to including these applications in the analysis.

Forty-one duplicate applications were identified during endline planning activities, and an additional 431 were identified during the endline. These are dropped from the analysis sample as well.[24]

Applicants to the scholarship program come predominantly from lower income households.

[23] Only 52 of the 683 applicants to this school were randomly assigned to receive a scholarship. The choice to drop these applicants from the follow-up was made because the additional observations would not meaningfully improve precision, but would require substantial additional data cost.

[24] The point estimate for the P_{priv} sample is 0.015 and is statistically significant. The point estimates in all other samples range from 0.002 to 0.006 and are not statistically significant.

Sixty-six percent of non-recipient households in the P_{main} sample report having a dirt or mud floor, 55 percent report mud walls, and 37 percent have access to electricity (Panel A of Table 2).[25] Across all samples, the most common reported occupation of the primary parent is casual labor, although a large share report working in agriculture or owning small businesses. Approximately half of the pupils are male. Overall, scholarship recipients and non-recipients are comparable in terms of baseline characteristics. Table 2 reports the F-statistic for the joint hypothesis that non-recipients and recipients are equivalent in terms of baseline covariates. For all sub-samples, the p-value for the F-statistic is greater than 0.05, providing some reassurance that scholarship awards were random within strata.

Panel B of Table 2 examines follow-up rates with the sample after applying all restrictions described above. The follow-up rate is above 85 percent for all sub-samples except for P_{priv}. Scholarship recipients are more likely to be observed at endline, but this difference is small and only statistically significant for the P_{main} and P_{priv} samples. Lee bounds for the main estimates are reported in Appendix Table A4.

3.5 Assessment design and construction of outcomes

At endline, pupils in pre-primary were tested in language skills (both Kiswahili and English) and math skills and primary school pupils were tested in English, Kiswahili, mathematics, social studies, and science. These subjects reflect the same academic subjects assessed in the KCPE exam, with the exception of religious education.[26]

The academic subject knowledge tests were designed by consultants with Kenya-specific teaching and curriculum development skills and experience, and were based on the Kenyan national curriculum. As such, although we do not observe to what degree Bridge schools' lessons focus on the Kenyan national curriculum on a day-to-day basis, we are able to observe how attending Bridge influences performance on tests based on the curriculum. Pupils were administered assessments corresponding to the grade they were finishing at the end of the 2017 academic year. For pupils in Standard 3 and below, the tests were administered as one-on-one interviews with field officers. For

[25] These figures reflect levels of socioeconomic status similar to the average in the Kenya. The 2014 Demographic and Health Survey shows that 64 percent of Kenyans do not have electricity, and 46 percent have cement floors (DHS, 2014).

[26] Religious education was omitted because of the complexity of developing and administering assessments for multiple religious groups.

pupils in Standard 4 and above, the tests were administered as pen-and-paper tests invigilated by a field officer.

All test scores are placed on a common scale using Item Response Theory (IRT) test equating procedures. Test score equating is important because the format (one-on-one or pen-and-paper) and difficulty of the assessment that pupils took could potentially be influenced by the scholarship offer– for instance, grade progression could differ between scholarship recipients and non-recipients. Tests for adjacent grades contained overlapping questions, allowing for comparisons across grades. For each subject test, a two-parameter IRT model is estimated and equating parameters between tests are estimated using the Stocking-Lord method.

We report our results in equivalent years of schooling (EYS) (Evans and Yuan, 2019) that scales all scores by the cross-sectional relationship between the test score and pupil grade.[27] The magnitude of effects in EYS units can be interpreted as the number of additional years of schooling that a pupil would need in a counterfactual school to learn as much as a pupil at a Bridge school during the study period. The PAPSA (Gray-Lobe et al., 2020) does not discuss this unit as EYS was an emerging concept and not very popular at the time the PAPSA was drafted. However, the use of EYS only rescales the point estimates and has no bearing on statistical inference.

We also present the results in Standard 4 standard deviation units, the unit originally specified in the PAPSA. Standard 4 was selected to serve as the base grade because it was the middle assessment.[28][29] The choice of units for the test scores affects the numerical values of the estimated effects of enrolling at Bridge. For example, it is possible to standardize the equated test scores to have unit standard deviation across the entire sample of pupils from pre-primary to through primary school grades. A standard deviation of this distribution is a much larger unit than a standard deviation of pupils within a single grade. Because few studies contain a single scale for pupils across such a large range of grades, expressing scores in units of the distribution of scores across all grades would have complicated comparisons to the existing literature.[30]

[27] EYS is closely related to the concept of learning adjusted years of schooling (LAYS), which rescales EYS units by a factor representing the relative effectiveness of a year of schooling in our context to that of a high performance benchmark (Angrist et al., 2020). We have not converted our estimates into those of the high performance benchmark because Kenyan EYS units seem suitable when the outcome is knowledge of material in the Kenyan national curriculum.

[28] Note that ECD grades received a single common assessment so there were four assessments for grades younger than Standard 4 and four assessments for grades older than Standard 4.

[29] Test equating procedures translate scores into a "base" assessment (Kolen and Brennan, 2014, Chapter 6).

[30] In the Online Appendix, we estimate effects on test scores expressed in several alternative units.

The pre-primary endline assessment included math, English and Kiswahili but not science and social studies. Analysis of the effect of enrolling at Bridge on science and social studies scores for the pre-primary sample is therefore restricted to the oldest pre-primary cohort – those who were entering the final year of pre-primary (or preunit in the Kenyan system) – because most of these applicants were enrolled in the first year of primary school at the time of the endline survey.

An *aggregate subject knowledge index* is constructed as the mean of standardized subject scores in math, English, and Kiswahili (Kling et al., 2007). The index is formed by first combining Kiswahili and English into a single *language index*. Indices are constructed using test scores expressed in standard deviation units of the Standard 4. The indices are then re-standardized to be expressed in units of the Standard 4 non-recipient distribution in order.

Additional indices are constructed to estimate the effect of Bridge on composite outcomes, including local content knowledge, higher order skills, instructional adaptation, classroom crowding, teacher-classroom engagement, test preparation, and parental engagement. See Appendix Table A2 for details on the construction of these indices. Some indices were not specified in Gray-Lobe et al. (2020). After observing similar effects for multiple thematically-related outcomes, we chose to include some new indices. In all cases, thematic relationships between items that were combined into indices were discussed in the analysis plan. Effects on the individual pre-specified outcomes that compose the indices are reported in the Online Appendix.

4 Empirical Framework

Because scholarships were randomly assigned, a comparison of outcomes between recipients and non-recipients identifies the average effect of *offering* a scholarship, the intent-to-treat (ITT) effect.[31] We begin by analyzing the effect of winning the scholarship on outcomes, also referred to as the intent- to- treat (ITT) or the reduced form estimate.

Estimation of the effect of enrolling at Bridge on outcomes requires additional assumptions. As will be discussed in Section 5.1, the scholarship doubled or tripled rates of Bridge enrollment for groups in our main analysis sample, but it also reduced the amount that households had to spend on their children's education out of pocket. By giving households a discount on the cost

[31] This section describes the empirical framework in general terms. A more detailed exposition, addressing technical details of the construction of some measures can be found in the PAPSA (Gray-Lobe et al., 2020).

of education, the scholarship may have increased household disposable income. This income effect may, in turn, have had important effects on outcomes of interest, such as test scores. Assuming that the scholarship affects test scores by increasing the chance that lottery winners enroll in Bridge and that other channels of impact can be treated as negligible, (effect of the scholarship on outcomes) = (effect of the scholarship on going to Bridge) $\times$ (effect of enrolling in Bridge). We measure the first two components of this equation by comparing scholarship recipients to non-recipients, and from there estimate the effect of enrolling in Bridge and from there estimate the effect of enrolling Bridge by dividing: (the effect of enrolling in Bridge) = (effect of scholarship on outcome)/(effect of scholarship on going to Bridge). Formally, this is done using two-stage least squares (2SLS). The ITT estimates identify the effect of receiving a scholarship on student outcomes, while the 2SLS estimates identify the effect of actually going to a Bridge school by accounting for exactly how much the scholarship increased Bridge attendance. In our case, the 2SLS estimates are more policy relevant and are therefore the focus of our study. The scholarship increased the probability of going to Bridge schools by one-third (Table 3).[32] Not all scholarship recipients chose to enroll at Bridge, and some non-recipients enrolled anyway. If Bridge enrollment is the only channel through which the scholarship affects test scores, the ITT estimate will be one-third the effect of enrolling at Bridge. Imbens and Angrist (1994) show that 2SLS estimates will capture the effect of enrolling at Bridge for those induced to attend by the scholarship (i.e. the Local Average Treatment Effect or LATE) if (a) receiving a scholarship to Bridge does not reduce the chance anyone enrolls at Bridge ("monotonicity") and if (b) winning a scholarship only affects outcomes by increasing the chance of enrolling at Bridge ("exclusion"). An alternative possibility we explore is that if, for example, winning a scholarship allows those who would have attended Bridge anyway to increase their consumption of goods and services other than Bridge, and if this consumption positively affects outcomes, then 2SLS estimates could overstate the impact of Bridge enrollment on outcomes.

The assumption that the exclusion holds approximately in our setting is motivated by empirical results that will be presented in Section 7 as well as prior literature. First, in Section 7.1 we show that the scholarship reduced educational expenditures by only between 15 and 25 percent at

[32] In approximate terms, receiving a scholarship causes Bridge attendance rates to rise from around one sixth to one half in the main primary-school-age sample (rising by about a third, thereby tripling), while these rates rise from around one third to two thirds in the main pre-primary-school sample (again rising by about a third, but because more pupils in this sample would attend Bridge regardless of scholarship status, this is a doubling rather than a tripling).

the primary school and pre-primary levels respectively and that this effect comes entirely from the reduction in fees, so households do not appear to have used additional income to increase non-school educational expenditures. Second, we show that the effect of the scholarship on test scores is small and insignificant in sub-populations where the first stage impact on Bridge enrollment increases linearly to the 2SLS estimate for sub-populations with larger first stage impacts. Meanwhile, the scholarship impact on the amount of fees paid is not predictive of the scholarship's impact on test scores. These results are consistent with the literature on the effects of cash transfers to households on educational outcomes. In Kenya, Haushofer and Shapiro (2016) show that a cash transfer worth over sixty times the value of the estimated income effect of the scholarship had at most a marginal impact on educational outcomes of children in the household.[33]

Randomization strata, as well as a wider set of baseline characteristics, allow us to explore the relative contributions of Bridge enrollment and possible income effects. Some strata – for example those who were enrolled in Bridge prior to the scholarship offer – have very high probabilities of attending Bridge even without the scholarship, but in this group the scholarship produces smaller first-stage enrollment changes (meaning any learning gain would be driven by income effects). Other strata have very low probabilities of attending Bridge without the scholarship, but large first-stage enrollment effects (meaning any learning gain would be driven by the shift to a Bridge-provided education). If the income effect were important, we would see large effects for the first group. However, we find the opposite: for pupils who were more likely to attend Bridge regardless of scholarship status, for whom any income effect would be most important, the scholarship had a smaller (and in some cases statistically insignificant) effect on test scores.

We extend the 2SLS analysis by including two endogenous variables in the empirical specification– enrolling at Bridge and the amount of school fees paid. This approach allows us to separately identify the effects of enrolling at Bridge and of paying fees, allowing us to test for an income effect under additional assumptions. We construct multiple instruments from interactions of an indicator for scholarship assignment and baseline covariates. The new instruments are used to

[33]The absence of large test score effects of cash transfers may reflect the fact that extra disposable income is likely spread across other priorities beyond the education of the scholarship recipient (Behrman and Parker, 2013). We report results in the Online Appendix that show that the scholarship raised the probability that siblings of scholarship recipients enrolled at Bridge. As an additional cross-check, we use the experimental variation that arose due to sibling scholarship receipt as an instrument for Bridge enrollment among scholarship non-recipients; we confirm that this produces similar point estimates for the impact of Bridge enrollment, although the much smaller set first stage means that this coefficient is not statistically significant.

estimate overidentified 2SLS models with multiple endogenous variables (Kline and Walters, 2016; Kirkeboen et al., 2016; Reardon and Raudenbush, 2013). Separate identification of the effect of fee payments and enrollment at Bridge comes from heterogeneity across applicants with different baseline covariates in the first stage effect on each endogenous variable. For example, pupils previously enrolled in Bridge prior to the scholarship experience much smaller enrollment shifts when they receive the scholarship, compared to those not previously enrolled in Bridge; however, both groups benefit from the reduced need to pay fees. Identification in this case works on the assumption that the average LATE is not predicted by the baseline covariates used to form the interaction terms (Hull, 2018; Reardon and Raudenbush, 2013). This assumption may be strong in some cases, and we discuss its plausibility for individual results.[34] Because these assumptions may not be equally plausible for all sets of baseline covariates, we evaluate robustness of the results by using different sets of covariates. The first set is a group of ten baseline covariates indicating the type of school that the pupil planned to attend in the absence of scholarship (public, non-Bridge private, or Bridge school), type of school the pupil attended at the time of the scholarship, socio-economic status, and whether the applicant applied through IPA. The second set of covariates is composed of predictions of the likelihood that a pupil would have, in the absence of the scholarship, enrolled in each of the three education options (public, Bridge, non-Bridge private) given their randomization stratum. These three predictions are calculated by first taking the mean non-recipient school type enrollment for the stratum. To avoid identification concerns that may arise due to overfitting of the predictions to the endogenous behavior of the control group (Abadie et al., 2018; Dean and Jayachandran, 2019), we ensure that all means leave individual i out of their own predicted choice. Means are shrunk toward overall mean choice using an empirical Bayes procedure which weights means by their reliability to improve predictive accuracy.[35][36] This procedure is applied to obtain separate predictions for each of the *main*, *brig*, and *priv* sub-samples. The Online Appendix includes robustness checks using additional sets of instruments, including interactions with county dummies.

[34] See Gray-Lobe et al. (2020) for a formal discussion.

[35] Simulations evidence obtained prior to the analysis of test score effects suggested that these adjustments were important. Leaving individual i in their own prediction led to an increase in type-I error. Empirical Bayes adjustment led to improvements in conventional measures of instrument strength including the Olea-Pflueger F-statistic and the Angrist-Pischke partial F-statistic.

[36] We model the means as draws from a binomial distribution and a beta prior. Details are provided in the PASPA.

We use a similar strategy to separately identify the effect of Bridge compared to public schools and the effect of Bridge compared to other private schools. Assuming income effects are negligible, if we use all sub-samples in our analysis, the 2SLS effect identifies the effect of enrolling at Bridge compared to a mixed counterfactual of alternative schools where compliers would have enrolled without the scholarship, including both public and non-Bridge private schools. We estimate the relative effect of enrolling at a public school and non-Bridge private schools compared to enrolling at Bridge using the same multiple endogenous variable 2SLS specifications used to separately identify the income effect and Bridge enrollment effect. As above, separate identification is valid with the assumption that the LATE is homogeneous across applicants for whom the baseline covariates used to construct the instruments are different.

We apply the 2SLS framework to evaluate effects on a range of pupil skills. While academic test scores measure an important component skill – one that is intrinsically important to parents who want to see their children excel in the school-leaving exams – it is important to consider whether the Bridge model supports the development of a broader range of skills that may also be important. One potential concern is that the detailed lesson plans may emphasize knowledge of curriculum over other cognitive and non-cognitive skills. A model that is very successful at getting pupils to produce the right answer to a test question may even impede the development of divergent thinking skills; a variety of non-cognitive skills may be important aspects of schooling experiences (Jackson, 2018). These concerns motivate an examination of the effect of enrolling at Bridge on alternative cognitive and non-cognitive assessments. We evaluate the effect of enrolling at Bridge using the same 2SLS framework described above on Raven's matrices, receptive vocabulary, executive function and digit span recall. We also test whether Bridge reduces creativity by estimating the 2SLS effect on an alternative use task measuring divergent thinking.

5 Impact on subject matter knowledge, grade progression, and other pupil outcomes

This section reports effects on enrollment (Subsection 5.1) subject-matter knowledge (Subsection 5.2), grade progression (Subsection 5.3), primary school completion (Subsection 5.4), and non-academic cognitive outcomes (Subsection 5.5) and non-cognitive outcomes (Subsection 5.6) among

the main sample (P_{main} and PP_{main}).

5.1 Effect of winning a scholarship on enrolling in Bridge and other school types

The scholarship increased the probability that pupils in the P_{main} sample enrolled in Bridge by 35.3 percentage points and pupils in the PP_{main} sample by 34.0 percentage points. (Panel A of Table 3). In the P_{main} sample, the scholarship reduced the share of pupils enrolling at public and non-Bridge private schools by 31.6 and 3.6 percentage points respectively, so the scholarship overwhelmingly moved children to Bridge from public rather than private schools. At the pre-primary level in the PP_{main} sample, the scholarship reduced the share of pupils enrolling at public and non-Bridge private schools by 14.3 and 18.9 percentage points respectively. Participation in formal education in 2017 is nearly universal for all samples, so few children who went to Bridge due to the scholarship would have been out of school had they not received a scholarship. As we discuss below, this has important implications for interpreting our results.

Although some pupils changed school types during the study period, the effect of the scholarship on years of exposure to Bridge is approximately twice the effect shown in Table 3, Panel A (0.7 years). Panels B and C of Table 3 examine the effect of the scholarship on enrollment in year 1 (2016) and on the number of years the pupil was enrolled at each school type across both years.

5.2 Effects on curricular subject knowledge

In this subsection, we show that Bridge enrollment dramatically increased scores on tests of subject matter knowledge (Subsection 5.2.1). Impacts are similar across baseline pupil and household characteristics (Subsection 5.2.2) and across Bridge locations (Subsection 5.2.3), consistent with an important role of standardization.

5.2.1 Average effects

The scholarship increased performance on all measures of subject knowledge for the primary (P_{main} sample) and pre-primary samples (PP_{main}). In primary schools, pupils gain 2.89 equivalent years of schooling after being enrolled at a Bridge school for two years compared to other schools. The effects are larger in pre-primary schools, where pupils gain 3.48 equivalent years of schooling in

pre-primary grades after being enrolled at a Bridge school for two years. These effects (shown in Table 4) are statistically significant at the one percent level.

Table 4 also reports the EYS gradient, the ordinary least squares (OLS) relationship between the test score outcome projected 2016 grade. This coefficient is the factor that is used to convert the test score from standard deviation units into EYS units. It is estimated on the sample of scholarship non-recipients who enrolled in public school. It should be noted that had we estimated the EYS gradient separately for the pre-primary and primary samples, the small non-linearities in the relationship between grade and performance on the tests would lead to slightly smaller EYS estimates for pre-primary and larger estimates for primary.

In standard deviation units, the ITT effect of receiving the scholarship on the aggregate subject knowledge index is 0.29 standard deviations and the 2SLS effect of enrolling at Bridge for two years is 0.81 standard deviations.[37] For the pre-primary sample, the ITT effect of receiving the scholarship on the aggregate subject knowledge index is 0.47 standard deviations and the 2SLS effect of enrolling at Bridge for two years is 1.35 standard deviations. In Appendix table A3, we show effects of enrolling at Bridge on specific items that were included on assessments taken by pupils in the pre-primary sample. For example, Bridge enrollment increases the number of letters identified per minute by 11.90, an 75 percent increase over the non-recipient mean of 15.8. Bridge increases the share of children who can count eight stars by 13 percentage points. Enrolling at Bridge also increases more advanced early reading and numeracy outcomes.[38] The probability of being able to read a simple sentence more than doubles. The effect on reading an English sentence is 30 percentage points, relative to a non-recipient mean of 22 percentage points; the effect on reading a Kiswahili sentence is 16 percentage points relative to a non-recipient mean of 11 percentage points. The probability of being able to do a simple addition problem rises 19 percentage points relative to a non-recipient mean of 50 percentage points.

We do not see evidence that Bridge is teaching to the test at the expense of higher order skills. Indeed, the effects are similar for the higher order skills index, an index based on the subject knowledge assessments that excludes items that may reflect recalling information or rote

[37] As discussed above, when we refer to standard deviation units as a measure of test scores, these refer to the dispersion of the Standard 4 population.

[38] It should be noted the pre-primary curriculum only requires that pupils be able to recognize letters and count to ten. More advanced numeracy and literacy skills may reflect the fact that Bridge increases timely grade progression and, as a result, the likelihood that a child has received instruction in these more advanced subjects.

memorization (Table A2). For primary pupils, the effect on the higher order skills index is 0.80 EYS and for pre-primary pupils, the effect on higher order skills is 0.82 EYS, statistically significant at the one percent level.

The point estimate for the effect of enrolling at Bridge on test scores is positive and statistically significant across individual academic subjects (Table 4). In primary schools, enrollment at Bridge increases scores on the aggregate science and social studies score by 1.35 EYS (Panel B of Table 4). The effects on science and social studies separately are similar. We also find that enrolling at Bridge increases scores on the local content material by 1.45 EYS.

5.2.2 Distributional and heterogeneous effects

The effect of enrolling at Bridge is larger at the bottom of the test score distribution than at the top. We compare effects at the bottom and the top quantiles using the Quantile Treatment Effect (QTE) estimator of Abadie et al. (2002), which estimates the LATE at different points of the distribution of test scores on those who switched to Bridge due to the scholarship (compliers). In primary school grades, the estimated quantile treatment effect on the 10th percentile of the distribution of compliers is 0.97 EYS, compared to 0.74 at the 90th percentile (Panel A of Table 5). The difference between the 90th and 10th percentile effects is statistically significant at the five percent level. In pre-primary school grades, the estimated quantile treatment effect on the 10th percentile of the distribution of compliers is 1.72, compared to 1.15 at the 90th percentile.

Another indication that Bridge compresses the test score distribution is that, in both the primary and pre-primary samples, the standard deviation of test scores is smaller among scholarship recipients (Panel C of Table 5). An F-test of the equality of variances in the scholarship and non-scholarship samples rejects the null hypothesis at the five percent level in primary school grades and at the 10 percent level in pre-primary grades.

The entire distribution of test scores is shifted to the right by Bridge enrollment at both the primary and pre-primary levels. Figure 6 shows the estimated empirical distribution for scholarship recipients and non-recipients for whom the scholarship would have determined whether they enrolled in a Bridge school (compliers). For both the primary and pre-primary levels, an empirical test that any point in the distribution of scholarship recipients is equal to the distribution of non-recipients is rejected at less than the one percent level indicating that, for compliers, the distribution of test

scores of scholarship recipients first order stochastically dominates that of non-recipients (Panel B of Table 5).

The compression of the test score distribution reflects, in part, larger effects for lower performing pupils who would have scored low on the endline test if they had not enrolled at Bridge. To quantify this a different way, we test for heterogeneity in the effect of enrolling at Bridge for pupils with varying baseline achievement. The data include several baseline test scores that are correlated with endline scores. Applicants enrolled by IPA officers have scores on Uwezo numeracy and literacy assessments, Raven's matrices, MDAT, and head-knees tasks. Applicants enrolled through UWR have percentage scores in English and math on their previous term 3 endterm exams. One option would be to separately estimate interactions with all of these baseline test scores. However, this approach raises two issues. First, all of these scores potentially measure other skills, beyond academic subject knowledge. Our goal is to test whether the effect is larger for those pupils who would have scored lower on the endline assessments if they had not enrolled at Bridge. Therefore, interaction effects with each individual test score may produce ambiguous results. Second, each specification would use a small portion of the data, resulting in many, less precise estimates.

We use baseline characteristics to obtain predictions of endline test scores and then use the prediction to test whether those expected to score lower on the test have larger effects from enrolling at Bridge. We obtain estimates of each pupil's *projected* endline test scores by fitting the model to the non-recipient sample. In the control group, predictions leave individual i out of their own prediction.[39] In primary school grades, the interaction term with the measure of expected achievement is -0.36 and statistically significant at the one percent level (Panel A of Table 8), suggesting that Bridge is particularly effective for pupils who started out most likely to do badly on the test at endline. In pre-primary grades, the point estimate of the interaction is negative but not statistically significant.

[39] Abadie et al. (2018) show that if individual i is left in the prediction, this can result in bias due to mean reversion in the treatment group. Baseline controls used to construct the projected endline test score are the self-reported term 3 end-of-term English and math scores prior to randomization, scores on individual cognitive and non-cognitive assessments, pupil height, pupil gender, a variable indicating whether the primary caregiver is literate, and primary caregiver occupation indicators (agricultural labor, casual labor, small business owner, employment status). Where data are missing for individual pupils, we impute the missing variable to the mean and include a dummy variable indicating that the value was missing. All predictions are formed separately within baseline grade cells and demeaned to have zero within baseline grade.

Heterogeneity across grades. The effect of enrolling at Bridge on academic test scores is over 1 EYS in pre-primary and the first year of primary school (Standard 1). This effect falls below 1 EYS starting in Standard 3 (Figure 7). The point estimates are between 0.5 and 1 EYS between Standard 4 and Standard 7. One explanation, consistent with the previous subsection documenting larger effects at the bottom of the test score distribution, is that the Bridge model is more effective at delivering less advanced content of lower grade instruction. However, three additional interpretations are also possible. First, tests at lower grades may measure skills improved by the Bridge model more accurately. It is, arguably, easier to measure (as well as develop scripted lesson plans for) basic numeracy and literacy skills like counting, letter and number identification, and the ability to sound out words may be easier to measure than more advanced math and reading skills. The effect of Bridge on the higher order skills index does not exhibit the same pattern across grades as the aggregate subject knowledge index (Figure 7). The higher order skills index removes items that reflect memory such as letter and number identification in early grades, suggesting that the difference is largely attributable to the prevalence of these types of tasks in the pre-primary assessments (Figure 7). Second, the effect of Bridge on mediating inputs (see Section 6) may differ across grades. The larger effects on subject knowledge in early grades correspond closely to the effect of enrolling at Bridge on the pupils' reported length of school day (Figure 7). In pre-primary grades, the effect of enrolling at Bridge is to increase the length of the school day by around 2 hours, in lower primary grades, there are smaller effects between 0.5 and 2 hours, and in upper primary grades, the effect on the length of the school day is approximately zero (Figure 7). It should be noted that in pre-primary grades, pupils are more likely to enroll at private schools in the absence of the scholarship. While the relative quality of these programs compared to public programs is unclear, the fact that the counterfactual is very different between grades should be considered in any comparison of the effects between pre-primary and primary grades. Finally, we note that the difference in the effects could also reflect the fact that the subject knowledge tests were administered as pen-and-paper tests after Standard 4 and as one-on-one interviews for pupils in Standard 3 and below.

Heterogeneity across household and location characteristics. Bridge increases test scores for households across a wide range of observable characteristics. We test for heterogeneity across

household characteristics such as household income (winsorized at the 99th percentile) and whether the household has a dirt floor and electricity. We also test for heterogeneity across characteristics of the location of the Bridge academy where the pupil applied, including the average KCPE (the exit exam) of schools within 2 kilometers of the Bridge academy, and whether the academy is located in a rural area. None of the estimated interaction effects are statistically significant at the five percent level. At the primary school level, the effect of Bridge may be 0.51 EYS smaller in rural locations (compared to 1.15 EYS in urban and peri-urban locations), however, the difference is only marginally significant.[40]

Heterogeneity by Bridge PTR There is no evidence of heterogeneity in the treatment Bridge effect by Bridge PTRs. For this analysis PTR is calculated as the number of pupils in a grade during the first term of 2017 according to internal test score data provided by Bridge.[41] To estimate this effect, we form an instrument by interacting scholarship assignment with the number of pupils in the grade the applicant was projected to be in at baseline.

At the primary level, our point estimate suggests that one additional student in a Bridge classroom is associated with a decline of 0.008 EYS in the treatment effect, but this estimate is not statistically significant (Table 8). While this lack of association does not prove lack of a causal effect, this finding is consistent with other literature that finds zero or small effects of PTR in developing countries (Duflo et al., 2015b; Banerjee et al., 2007; Kremer et al., 2013; Kremer and Holla, 2009; Banerjee et al., 2005). At the pre-primary level, the point estimate suggests that one student increase in class size is associated with a gain of 0.003 EYS, but this is not statistically significant (Table 8). This provides suggestive evidence that the Bridge effect is not primarily due to the smaller PTR.[42]

Heterogeneity by Teacher characteristics. We examine whether the effect of Bridge is mediated by several observable teacher characteristics that have been found to be predictive of teacher

[40] The rural designation comes from Bridge data. Academies are classified as either urban, peri-urban, or rural.

[41] At the time of this study, all Bridge schools had a single stream in each grade, and therefore the PTR coincides with class size.

[42] A back-of-the-envelope calculation using this evidence suggests that class size would explain less than one-fifth of the effect of Bridge. Even using the upper bound of the 95% confidence interval in (Duflo et al., 2015b) and scaling it for 20 pupils (the approximate difference in class size between the average public classroom and the average Bridge classroom), suggests that increasing class sizes by 20 pupils would reduce test scores by 0.14 standard deviations.

effects in other settings. We conduct these tests at the subject level because teacher characteristics vary between subjects. Our goal is to estimate the effect of enrolling at Bridge interacted with the characteristic of the teacher assigned to the grade where the pupil enrolled. To estimate this effect, we form an instrument that interacts scholarship assignment with the characteristic of the teacher in the grade that the applicant was projected to be in at baseline.

The teacher characteristics we study here are an indicator for whether a teacher has more than 2 years of experience teaching, the number of years the teacher has taught at Bridge, and the teacher's performance on the Kenya Certificate of Secondary Education (KCSE). All teacher characteristics data come from administrative files maintained by Bridge. The date of hire and KCSE score are recorded in recruitment files. Bridge teacher data are matched to applicants using a file that links Bridge teachers to grades and subjects taught in the first term of 2017. The file only contains primary school grades, so we restrict the analysis to the primary school sample.[43]

The KCSE is graded on an alphabetic scale with A+ being the highest grade, and E being the lowest. Grades A through D have intermediate scores indicated by '+' and '-'. Alphabetic scores are transformed to a numeric scale by awarding a single point for each letter above E, and then adjusting by 0.2 or -0.2 for plus or minus respectively. Over half of teachers received a C and 35% received better than a C. For all results that follow, we standardize the numerical KCSE score to have mean zero and unit standard deviation across the applicant sample.

Years experience is calculated as the number of years that had elapsed at the beginning of the 2016 academic year (year 2 of the program) since the teacher was hired.[44]

The Bridge effect appears to be unrelated to observed characteristics of teachers in Bridge schools. In primary school, the effect of Bridge on test scores when teachers have less than two years of teaching at Bridge is 0.73 EYS. We are also unable to reject the null hypothesis that the effect of enrolling at Bridge is unrelated to the teachers total number of years of experience or the teachers KCSE score. This may be especially surprising given that multiple studies have demonstrated that novice teachers are less effective (Chetty et al., 2014; Bau and Das, 2021) and

[43]Technically, we can also estimate these effects for those applicants in the oldest cohort of the pre-primary sample. However, the characteristic of teachers in the grade that the pupil actually reached is only observed if the teacher reached primary school. Because enrollment at Bridge affected grade progression, estimates from this specification are potentially compromised by selective observation of teacher characteristics.

[44]This may understate total number of years experience, but other studies documenting the impacts of years of experience have the same limitation (Chetty et al., 2014).

that teacher knowledge with higher levels of subject knowledge are more effective (Bau and Das, 2021; Bold et al., 2019).[45]

5.2.3 Standardization and uniformity of output

Two tests suggest Bridge's attempts to codify and standardize education produced more uniformity. We first show that we cannot reject the hypothesis that the effect of enrolling at Bridge is equal across Bridge locations (sites) using an overidentification test that instruments Bridge enrollment in the second year of the scholarship program with interactions of treatment with location dummies.[46] The null hypothesis of the over-identification test is that the effect of enrolling at Bridge is equal at all academies (Walters, 2015). We also estimate the variance of academy effects directly using a procedure from Walters (2015), finding less dispersion among Bridge academies than among others This procedure estimates the variance in effects across sites under relatively strong parametric assumptions about the distribution of parameters which we describe below.

Using the overidentification test, we are unable to reject the null hypothesis of no heterogeneity in the effect of Bridge enrollment across locations at either the primary (Column 1, Panel A of Table 6) or pre-primary (Column 2) level. This result is consistent with the view that the effect of Bridge is similarly positive at all of its locations. In Gray-Lobe et al. (2020), we had pre-specified that this overidentification test would serve as our primary statistical test of standardization across Bridge schools. However, this result may also indicate a lack the statistical power to detect variation in the effect of different Bridge schools due to the small number of observations in each location. [47] We therefore complement this analysis with the procedure from Walters (2015).[48]

We estimate the dispersion in Bridge effects directly under stronger parametric assumptions

[45] In the Online Appendix, we report more precise estimates using a non-experimental, pupil fixed effect estimator similar to Bold et al. (2019). This non-experimental strategy yields standard errors that are less than one third the size of the interaction effects. Still, in all cases, we are unable to reject the null hypothesis, suggesting that the Bridge is effective even when teachers are inexperienced or have low subject knowledge.

[46] Some readers may be familiar with the use of overidentification to test the null hypothesis that all instruments are valid under an assumption of constant effects. Because our instruments are constructed using randomly assigned scholarship offers, we are confident in assuming their validity. The overidentification test is therefore of the null hypothesis of effect homogeneity (Angrist and Pischke, 2009, p. 167).

[47] We find similar results pooling the primary and pre-primary samples to obtain a larger sample size (Column 3 of Table 6).

[48] This procedure was prespecified in Gray-Lobe et al. (2020) as providing suggestive evidence on the degree of standardization afforded by the Bridge model. However, at the time, we planned only to apply (Walters, 2015) in the event that we rejected the null hypothesis of the overidentification test above. We decided that these results would be informative regardless of the result of the overidentification test.

(Panel B of Table 6) using the procedure from Walters (2015). We model potential outcomes of pupil i in location j using a random coefficients specification

$$Y_{ij}(d) = \alpha_{dj} + \varepsilon_{idj}$$

where $d \in \{0, 1\}$ indicates Bridge enrollment and $E[\varepsilon_{idj}] = 0$. The decision to enroll in Bridge is described by a "Heckit" or threshold crossing model (Vytlacil, 2002; Heckman, 1976):

$$D_i = \mathbb{1}\left\{\lambda_j + \pi_j Z_{ij} > \eta_{ij}\right\}$$

Each location is therefore characterized by

$$\theta_j = (\alpha_{1j}, \alpha_{0j}, \lambda_j, log\pi_j)'.$$

The term $log\pi_j$ ensures that a Bridge scholarship offers always increases the probability of Bridge enrollment. The error terms are assumed to be distributed multivariate normal:

$$(\varepsilon_{i1j}, \varepsilon_{i0j}, \eta_{ij})' \sim N(0, \Sigma).$$

Walters (2015) shows that under these assumptions, we can derive the integrated likelihood function, which can then be estimated using the method of simulated moments.

We first conduct a validation exercise. Despite the parametric assumptions, the model estimates that the mean Bridge enrollment effect (akin to the treatment effect of Bridge schools) is 0.909 EYS compared to our 2SLS estimate of 0.89, providing us with some reassurance about its validity. At the primary school level,a standard deviation in the effect of Bridge on test scores across locations is 0.339 EYS. This means that the top decile of Bridge site effects is approximately, $1.65 \times 0.339 + 0.909 = 1.468$ EYS, whereas the bottom decile is approximately 0.350 EYS. The standard deviation of site effects at the pre-primary level is slightly smaller 0.282 EYS with a mean effect of 1.439 EYS. This means that the top decile of Bridge site effects is approximately 1.904 EYS, whereas the bottom decile effect is approximately 0.974 EYS. Taken together, the results suggest that even the least effective Bridge academies still have significant effects on test scores.

The dispersion in site effects depends both on the dispersion of the mean Bridge outcome, and the mean non-Bridge outcome. So far we have considered whether the difference in the complier means varies across sites. However, the Walters (2015) procedure also yields estimates of the variance in the complier means themselves. At the primary school level, we find that the standard deviation of the complier mean for scholarship recipients is 0.253, less than three quarters the non-recipient complier mean. At the primary school level, therefore, outcomes at Bridge schools are more uniformly distributed than counterfactual outcomes. At the pre-primary level, the standard deviation of the complier mean is 0.422, approximately 29 percent greater than the counterfactual mean, so that test scores at Bridge schools are less uniformly distributed than counterfactual outcomes. At the pre-primary level, these means are positively correlated, suggesting that some of the variation in Bridge outcomes may be due to unobserved characteristics, such as family background.

5.3 Effects on timely grade progression

Some schools might hold back pupils' grade progression so as to strategically improve test results in key grades. Focusing our analysis on the primary (P_{main} sample) and pre-primary samples (PP_{main}), we see no evidence that Bridge is more likely than public schools to hold back pupils. Rather, our results show that pupils are less likely to be held back at Bridge. On average, the endline data show that 74 percent of non-recipients in the main primary school sample progressed to higher grades in a timely fashion. Pupils who were enrolled at Bridge for two years increased the likelihood of timely grade progression by 20 percentage points (Table 7). In the main pre-primary sample, 74 percent non-recipients progressed to a higher grade in a timely fashion, and enrolling at Bridge increased the probability of this by 19 percentage points. These effects appear to have persisted two years after the end of the scholarship program. The point estimate for the effect of enrolling at Bridge on timely grade progression by 2019 is smaller, but still positive and statistically significant at the one percent level.

5.4 Effect on Primary School Completion

Enrolling in a Bridge school improves primary school completion outcomes, measured using self-reported KCPE test taking and KCPE scores. The KCPE is the eighth grade nationwide primary

school exit exam that serves as a credential for primary school completion and is required for entry into secondary school. Given the timing of our data collection, only the oldest three cohorts of pupils – those cohorts who were entering Standard 5 or higher at the start of the scholarship program— could potentially sit for the KCPE during our follow-up survey period. This limits our sample size and statistical power on outcomes related to the KCPE and the secondary school transition.

The timing of the scholarship relative to the KCPE test taking varies by cohort. Pupils in the oldest cohort (seventh grade at the start of the study), these pupils could receive scholarship support for the last two years of their primary education, including the year in which they took the KCPE. For the younger cohorts, the scholarship support would end before they took the KCPE, leaving these cohorts without any scholarship funding for one or more primary school years (including the year of KCPE taking).

Because of the scholarship timing differences (relative to the KCPE) across cohorts, we report results separately by cohort, but also present the pooled results to preserve statistical power. Due to the smaller sample sizes for these analyses these tests include a linear control for the probability of scholarship assignment in the randomization strata to avoid loss of sample size due to strata cells with no variation in the scholarship instrument.[49] We present the results in Table 9. Due to the analytical complications outlined above, these results should be interpreted as suggestive.

Enrolling at Bridge increases KCPE test taking rates by 15 percentage points (Table 9). This effect is largest and statistically significant for the younger two cohorts. Enrolling at Bridge increases the average test score for those pupils who take the KCPE by 16.2 points.[50] Enrolling at Bridge also increases the probability that a pupil passes the KCPE by 17 percentage points over a base of 41 percent. This effect, statistically significant at the one percent level, is in part due to the effect on test taking. The estimated effect of enrolling at Bridge on passing the KCPE conditional on taking the exam is 12 percentage points, but this effect is not statistically significant.

Because of the salience of the KCPE, one concern is that Bridge schools might discourage pupils who are likely to obtain low KCPE scores from taking the exam in an effort to artificially raise the average performance (and reputation) of their schools. However, across all three cohorts we

[49]The estimated coefficients using strata fixed effects are similar in magnitude, but are less precisely estimated.

[50]In the sample of non-recipients, a standard deviation of the (self-reported) KCPE test score distribution is 58 points.

find that attending Bridge increases the probability of taking the KCPE by 15 percentage points (Column 2). This effect is driven by the large positive effects among the younger two cohorts (Columns 3 and 4). However, enrolling at Bridge reduces KCPE taking by eight percentage points, however, this coefficient is imprecisely estimated (Column 5).

In terms of KCPE test scores, attending a Bridge school increases performance on the KCPE by 16.2 points (relative to a mean of 260), and increases the probability of passing by 17 percentage points. These results are potentially confounded by the changes in the composition (or selection) of test-takers. The treatment effect of Bridge on KCPE test scores varies inversely with the effect on KCPE test taking within each cohort (Columns 3-5). This result supports the interpretation that pupils who are induced to take the KCPE by attending Bridge tend to be lower performing and therefore the estimated 2SLS effect of enrolling at Bridge on KCPE scores may reflect a negative selection effect in addition to the average causal effect of enrolling at Bridge.

Pupils are recorded as transitioning to secondary school if they reported enrolling in secondary school at endline at the end of 2017 or in a 2019 phone call survey. The point estimate indicates that enrolling at Bridge increases the probability of a pupil attending secondary school by 2 percentage points over a base of 71 percent but this effect is not statistically significant (Table 9). The absence of a statistically significant effect on secondary school transitioning despite positive effects on grade progression may reflect the fact that financial constraints remain a major barrier to secondary school transitioning. Given the smaller sample size associated with this analysis, it is also likely that we do not have sufficient statistical power to detect small increases in secondary school enrollment.[51]

5.5 Effect on non-subject-matter cognitive outcomes

This section discusses the effect of the scholarship and Bridge enrollment on measures of general cognitive ability, working memory, receptive vocabulary, executive function, and divergent thinking.[52]

[51] Ozier (2018b) estimated the relationship between KCPE passing and secondary school enrollment in western Kenya. Our estimated secondary school effect is consistent with his estimates when they are scaled by our treatment effects of Bridge enrollment on KCPE passing.

[52] Unless otherwise indicated, all tasks are scored as the count of correct responses. Scores are standardized to the distribution of Standard 4 scholarship non-recipients, except where indicated otherwise.

Raven's progressive matrices. Raven's progressive matrices are commonly used to measure fluid intelligence, and have been previously used in Kenya (Dean et al., 2018; Ozier, 2018a,b; Friedman et al., 2016). Pupils were asked to identify which of six possible patterns fit with a given matrix of three patterns. Scholarship recipients score higher than non-recipients on Raven's matrix nonverbal reasoning tasks, but the differences are not statistically significant in either the pre-primary or primary samples (Table 10).

Forward digit span recall. The digit span recall task is a common measure of working memory, also previously used in Kenya (Dean et al., 2018; Ozier, 2018a). Pupils were asked to repeat strings of four to seven digits. The task may also measure numeracy skills if familiarity with the names of numbers facilitates recall. At the primary school level, pupils who enroll at Bridge gain 1.23 EYS compared to non-Bridge schools (Table 10). In pre-primary grades, the effect of enrolling at Bridge is 1.76 EYS compared to non-Bridge schools. In both cases, the effects are statistically significant at the one percent level.

Draw-a-person task. The draw-a-picture task is a measure of general early childhood cognitive development. The task is an abbreviated version of the Goodenough-Harris task (Goodenough, 1926; Harris, 1963). Pupils were asked to draw a single picture of a person. The score was calculated by assessing whether the drawing included seven characteristics: (1) a head, (2) a body, (3) arms, (4) legs, (5) hands or feet, (6) one recognizable facial feature, and (7) two recognizable facial features. This task was only administered for pupils in pre-primary grades at endline and the scores are standardized to the preunit (final pre-primary) grade. Enrolling at Bridge leads to a gain of 1.07 EYS, statistically significant at the one percent level (Table 10).

Receptive vocabulary task. 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, drawn from assessments developed in Kenya by Knauer et al. (2019). This task was administered to pupils in Standard

2 and below. The outcome is expressed in standard deviations of the Standard 2 distribution. For pupils in the primary school sample who did participate in this task, the point estimate of the effect of enrolling at Bridge is 0.31 EYS, but this is not statistically significant. For the pre-primary sample, the effect of enrolling at Bridge on this task is 1.21 EYS, which is statistically significant at the one percent level.

5.6 Effect on non-cognitive outcomes

Head-knees task. The head-knees task is a measure of inhibitory control based on McClelland et al. (2007). This task was administered only for pupils who were in pre-primary grades at endline. This task measures behavioral self-regulation. The task begins with a practice round in which the pupil is asked to touch their head and their knees. After that, the field officer asks the pupil to do the opposite of what they are asked. The field officer records that either the pupil responds correctly, incorrectly, or self-corrects. This task was only administered for pupils in pre-primary grades at endline and the scores are standardized to the preunit (final pre-primary) grade.[53] The effect of enrolling at Bridge is 0.64 EYS (Table 10), although the effect is only marginally statistically significant.

Spoon task. This task is adapted from alternate use tasks commonly used to measure creative ideation (Guilford, 1967; Torrance, 1972; Yando et al., 1979; Plucker, 1999). For this task, pupils are instructed to imagine that they have been given a spoon; they are then asked to list potential uses of the spoon.[54] The interviewer recorded each response and categorized the response as either "normal" or "unusual."[55] Field officers were instructed to keep track of the number of "different" responses that the pupil listed. Responses were considered to be "different" if the movement of the spoon was different.[56] Pupils were given two minutes to think up as many uses as they could.

[53] We score respondents who fail to complete a preliminary, introductory practice round successfully as scoring zero. We also score respondents who initially respond incorrectly, but self-correct as scoring zero.

[54] The exact wording was "Let's play another game: Imagine I gave you a spoon as a present. Can you tell me all the different things you can do with the spoon? Remember it is your spoon so you can do anything you want with it. There is no right or wrong answer. I just want to know all the different ways you can use a spoon. It can be at home, at school, with friends, in the shamba, anywhere. I just want to hear all the things you can do with a spoon."

[55] "Normal" uses were those for which "the spoon is designed for." Examples were given of eating, cooking, or stirring tea. "Unusual" uses were defined as those for which the spoon was not designed. Examples were given of digging, making music, or playing games.

[56] Eating was soup and eating rice were given as examples where the movement was the same. Different games were treated as different uses.

To ensure that grading is uniform across respondents, we clean the data using regular expressions to identify instances of the most common uses and then remove duplicate entries.[57] Grading instructions for this task were not uniformly followed in all interviews.[58] Inconsistencies were identified and corrected.[59]

Enrolling at Bridge has no discernible effect on performance on the spoon task. We report two scores derived from the results of this task. The first is the number of field officer recorded "unusual uses." This outcome is intended to be comparable to what is commonly referred to as "originality" in the evaluation of alternative use tasks (Torrance, 1972; Plucker, 1999). We also report the total number of uses enumerated by the pupil; we refer to the latter measure as ideational "fluency." Neither outcome is affected by enrolling at Bridge.

6 Effect on intermediate inputs and classroom experience

Pupil reports about their classroom and school characteristics suggest that Bridge's attempts to codify education has significant impacts on some, but not all aspects of pupils' day-to-day classroom experience.

Pupils in Bridge schools report more instructional time than in other schools, consistent with their longer teaching hours and codified monitoring processes. Pupil reports indicate that enrolling at Bridge increases a school day by 0.59 hours at the primary level and 1.79 hours at the pre-primary level (see Table 11). Enrolling at Bridge increases the likelihood of attending Saturday school by 40 percentage points at the primary school level, and 24 percentage points at the pre-primary level. Enrolling at Bridge increases the probability of never experiencing teacher absence from class by 14 percentage points. At the pre-primary level, enrolling at Bridge does not seem to affect reported teacher absence.

Pupils in Bridge are less likely to report the use of local languages for instruction, consistent with the exclusive use of English and Kiswahili in the centrally designed scripts. The effect is 19 percentage points at the primary level and 11 percentage points at the pre-primary level.

[57] These include eating, cooking, stirring, digging, writing, scooping, cutting, measuring, and smearing.

[58] Some field officers listed instances of the same use ("eating ugali", "eating vegetables"). According to the instructions, these uses should only have been counted once. In other cases field officers listed multiple uses as a single use ("cooking, eating, and stirring").

[59] It should be noted that this data cleaning was not specified in Gray-Lobe et al. (2020) and was conducted after initially analyzing the data. These data cleaning choices have a negligible impact on the estimated effects.

A concern with the use of teacher scripts in education and the use of standardized procedures in general is that it may prevent teachers from adapting instruction to the individual needs of pupils (Dresser, 2012; Valencia et al., 2006). However, both pupil and parents reports suggest that Bridge is more likely to provide pupils with instruction or materials related to their individual needs at the primary and pre-primary level (see Table 11). Bridge also promotes teacher and parent engagement in children's education. The estimate of teacher-classroom engagement index is positive but statistically significant at the five percent level only at the primary level (Table 11).[60] Enrolling at Bridge also increases the parental engagement index in both the samples.

Enrolling at Bridge increases reported homework and test practice. Enrolling at Bridge increases the number of days in a usual week that a pupil reports having homework by 1.7 days in the primary school sample and by one day in the pre-primary sample (Table 11). Enrolling at Bridge increases the point estimate of a test preparation index in both primary and pre-primary samples but the estimates are not statistically significant at the pre-primary level (Table 11).

Bridge pupils are also more likely to receive extra support from their teachers to prepare for tests. In the primary sample, enrolling at Bridge increased the probability of pupils reporting receiving help from their teachers to take a test like the one that is administered at endline by 9 percentage points (Table 11). This may be a concern if Bridge teaches to the test and only helps pupils develop test-taking skills. However, we have reasons to believe why this may not be the case. First, Bridge was not informed about the content of our assessments and if they taught similar content, it is largely due to the correlation between the content of our assessments and the curriculum. Second, we find no evidence that the test score results are driven by Bridge teaching to the test at the cost of higher order skills. We find positive and statistically significant effects on higher order skills. Third, it may be possible that Bridge pupils developed better test-taking strategies as they are tested up to seven times in each subject in an academic year. However, in a different setting, Berry et al. (2020) estimate very small impact of continuous evaluation, suggesting that Bridge effects are unlikely to be explained by greater exposure to test-taking.

Enrolling at Bridge also increases the number of textbooks that a pupil has by 0.65 books in primary grades and by 0.48 books in pre-primary grades. (Table 11).

Bridge has been less successful at improving pupil safety in schools. While Bridge schools are

[60] See Appendix Table A2 for details on the construction of all indices discussed here.

more likely to have a wall or a fence around school, pupil reports of hazards in the playing fields were slightly higher compared to other schools. Bridge reduces pupil-reported corporal punishment, but most pupils still report corporal punishment (Table 11).[61] Enrolling in a Bridge school reduces the fraction of pupils reporting that their teachers carry sticks or canes that can be used to hit pupils by 14 percentage points from a base of 81 percent. Eighty-three percent of non-recipient pupils in primary grades report teachers hitting, slapping, or pinching pupils, and enrolling in Bridge reduces the share reporting these behaviors by 6 percentage points.[62] In pre-primary grades, we do not find any statistically significant effect of enrolling at Bridge on corporal punishment.

7 Are effects driven by financial factors?

In this section we consider whether the scholarship may have had an effect on academic subject knowledge of pupils beyond its effect on Bridge enrollment, a central assumption needed to identify the effect of enrolling at Bridge on test scores described in Section 4. This section begins by describing some possible effects of the scholarship that would be inconsistent with the "exclusion restriction" that the scholarship only affected pupil outcomes through its effect on Bridge enrollment. In Section 7.1, we examine the impact of the scholarship on fees paid, tutoring expenditures, and missed classes – three intermediate outcomes that could influence pupil test scores beyond their impact on the probability that a pupil enrolls at Bridge. Finally, Section 7.2 discusses estimates of the separate effects of enrolling at Bridge and paying fees within a multiple endogenous variables framework.

The scholarship may have had an effect on households beyond its effect on Bridge enrollment. The scholarship covered tuition and uniform costs, potentially saving households money and helping families remain in good standing with school administrators. The scholarship may have reduced missed class due to an inability to pay fees. It may have freed up household money for additional tutoring, better food, medical care, or reduced stress in the pupil's home environment. Paying fees may also have a positive effect on pupil learning if they feel a duty to study more when their families

[61] As noted earlier, Bridge explicitly forbids corporal punishment (Kwauk and Robinson, 2016).

[62] Note that these student-reported levels based on recollection are much higher than, and not directly comparable to, the roughly 5 percent of school-level observations (authors' calculations) from the World Bank's Service Delivery Indicators in which teachers in government schools were actually seen hitting, pinching, or slapping pupils during an enumerator's single visit to the school (Martin and Pimhidzai, 2013).

are paying, if parents are more engaged because they are paying, or if schools focus attention on pupils who are "paying customers." Also, as shown in Section 6, Bridge schools operate longer school days and have more Saturday school, so it is possible that enrolling at Bridge could crowd out other educational expenditures such as private tutoring.

7.1 Effects on fee payment, tutoring, and missed class

In this section we examine the impact of the scholarship on household finances and related outcomes directly.[63] We show that while the scholarship reduced the amount of money households had to spend on education and the amount of school that recipients missed, the magnitude of the effects are unlikely to be driving the estimated effects. Furthermore, we find no evidence that households used the funds freed up from school fee payments due to the scholarship for other educational inputs such as extra tutoring.

The scholarship reduced the amount that households spent out of their own pockets on their pupils' education by between 15 and 30 percent. Families of non-recipient pupils in the primary school sample spent about KES 7,500 annually on education in the absence of the scholarship; scholarship assignment reduced this amount by KES 1,270 (Table 12). This effect appears to be driven primarily by a reduction in school and registration fees. Non-recipients in the primary school sample paid KES 3,690 on average in fees; the scholarship reduced this amount by KES 1,350. In the pre-primary sample, scholarship non-recipients paid approximately KES 8,740 annually on education expenditures, of which KES 5,420 was on fees. The scholarship reduced total education expenditures by KES 2,490 and reduced fees specifically by KES 2,320. These patterns are broadly consistent with an elimination of school fee payments among the roughly one-third of recipients who chose to make use of the scholarship offer.

The scholarship reduced the likelihood that pupils missed class due to an inability to pay fees at both the primary and pre-primary level (Table 12). At the primary school level, the scholarship reduced the share of caregivers who reported their child missing class due to an inability to pay fees by 19 percentage points from a base of 47 percent. The scholarship recipients missed 0.37 days fewer in the final term of 2017 at the primary school level and 0.53 days fewer at the pre-primary

[63] Analogous results for the PP_{brig} and P_{brig} samples are reported in Appendix Table A5. The results are broadly similar to those Table 12. These samples will be included in the next subsection.

level.

The point estimate of the effect of the scholarship on tutoring expenditures is negative, although it is not statistically significant in the primary school sample and only statistically significant at the 10 percent level in the pre-primary sample (Table 12). In the primary school sample, households spent about KES 530 annually on tutoring activities and about 31 percent of pupils received some tutoring. In the pre-primary sample, these figures are closer to zero: pre-primary households spent about KES 360 annually on tutoring activities and about 18 percent of pre-primary pupils received some tutoring. The scholarship reduced tutoring expenditures by about KES 60 in the primary sample and by KES 100 in the pre-primary sample.

7.2 Estimating the effect of paying fees

We extend the 2SLS framework described in Section 4 to estimate the effect of enrolling at Bridge and paying school and registration fees.[64] The scholarship reduced the fees that households had to pay to enroll their children in school and increased the likelihood that children enrolled at Bridge. We can consider the possibility that both of these are possible channels through which the scholarship may have affected test scores.

These two ways in which the scholarship affected recipients correspond to separate first stage impacts of the scholarship. The 2SLS framework described above treats the impact of the scholarship on test scores through the relaxation of the household budget constraint as small enough that it can be ignored. An alternative view could be that Bridge enrollment has a negligible impact on test scores. In this case, we could estimate directly the impact of paying fees in a similar 2SLS framework using the first stage impact of the scholarship on fee payment.

As a first step, we explore which of these two models better explains our data. To do so, we first stratify the sample by county and probable outside option school type, pooling all available samples (*main*, *brig*, and *priv*). The effect of the scholarship on fees and Bridge enrollment naturally varies across strata cells. We then examine how the ITT effect of the scholarship on test scores varies with the Bridge enrollment effect and the fee payment effect. This exercise is analogous to that of

[64] At the pre-primary level and in private primary schools, these costs would be predominantly tuition fees. Public primary schools are prohibited from charging tuition fees, but parents still report paying fees for school improvement and meals. The survey asked parents to report the total amount paid in the past year for "school and registration fees, including meals". Costs for uniforms, books, transportation, boarding, and extra tutoring sessions were collected separately.

visual-IV plots (Angrist and Pischke, 2009, Chapter 4) for each mediator.

The exercise shows that while the first stage effect on fees alone cannot explain the pattern of results, the first stage effect on Bridge enrollment can. Figure 8 shows the relationship between the ITT effect on test scores and the first stage effect on fees (top two panels) and Bridge enrollment (bottom two panels). Shaded circles show the ITT and first stage effect for each individual stratum. The size of the circle is inversely proportional to the standard error of the 2SLS estimate. The confidence interval of the best fit line through the circles is given in orange. The red line indicates the pattern that would be expected if the effect of the scholarship were mediated entirely by fees (top two figures). The green line in the bottom two figures indicates the pattern that would be expected if the effect of the scholarship were mediated entirely by Bridge enrollment. These lines pass through the origin with slope equal to the estimated 2SLS effect across all strata of fees (top panel) and Bridge enrollment (bottom panel). This slope, the just-identified 2SLS effect, is the ratio of the ITT effect on subject knowledge index (in Table 4) to the ITT effect for the associated first stage (in Table 12) for either fees or enrollment, respectively. The results in the top panel show that, contrary to what would be expected if income effects were responsible for the learning gains (shown by the red lines), the relationship between the income effects of the scholarship and ITT effects on learning is approximately flat. Even for strata cells where the scholarship is estimated to have reduced fees paid by large amounts, the scholarship's effect on test scores is approximately the same as it is for cells where the scholarship is estimated to have had virtually no effect on fees. In contrast,, the results in the bottom panel show that, consistent with what would be expected if enrollment in Bridge were responsible for the learning gains (shown by the green lines), there is a strong positive relationship between the first stage effect on Bridge enrollment and the ITT effects on learning. Strata cells where the scholarship is estimated to have had little effect on the decision to enroll at Bridge are associated with null ITT results.

An important fact illustrated by these results is that even in populations where the income effect is very small, the scholarship is approximately equally effective at improving test scores. Relatedly, we find very small scholarship impacts on test scores for strata cells where the scholarship had a small impact on Bridge enrollment.

These results tell us both that the scholarship's impact on test scores cannot be explained by the effect on fees alone and that at least for some households where the scholarship did not increase

fees, the 2SLS assumptions are justified.

To go further and directly estimate the effect of reducing the amount households pay in fees we expand the 2SLS model to include multiple endogenous variables: one for Bridge enrollment and the other for school and registration fees paid. Building off of the VIV results above, we will identify the model using variation in the first stage impact on different sub-populations on the two endogenous regressors. This strategy comes at the expense of an additional assumption that the effect of both paying more in fees and Bridge enrollment is the same for households with different values of the baseline characteristics used to identify the different populations. For example, if household wealth is predictive of larger first stage effects of the scholarship on fee payment (because these households tend to enroll their children in more expensive counterfactual schools), then if we assume the effects of paying additional school and registration fees and enrolling at Bridge are similar across more or less wealthy households, then the 2SLS model would be identified by the variation in the first stage impacts.

We would like to estimate the separate effects of having to pay more in fees and enrolling at Bridge to assess whether the 2SLS framework described above is appropriate. To formally estimate the separate effect of paying fees and enrolling at Bridge, we use an overidentified 2SLS specification using multiple instruments formed by interacting scholarship offer status with baseline covariates. The separate effects of these two mediators are identified under the assumption that the baseline covariates used to form the instruments are not predictive of the local average treatment effect for compliers (Hull, 2018). Because this assumption is relatively strong, we estimate specifications using two different sets of instruments to assess robustness. The first set includes a set of baseline characteristics related to demographics and variables that may be predictive of counterfactual enrollment, including stated enrollment plans: being enrolled in any school at time of application; being enrolled in NGO/private school; any prior Bridge enrollment; having plans to enroll in a Bridge school/public school/non-Bridge private school in the absence of the scholarship; primary caregiver is a agricultural/casual laborer; and whether the household applied through IPA. The instruments also include an inverse hyperbolic sine (IHS) transformation of reported household monthly income. The second set is meant to leverage variation in counterfactual mediators/counterfactuals at the Bridge location level. Including a full set of interactions with 405 Bridge location indicators risks overfitting due to many weak instruments (Andrews et al., 2019).

Similar to Dean and Jayachandran (2019), we collapse many variables into fewer scalar controls by estimating the leave-i-out non-recipient mean school enrollment state (Bridge, public, non-Bridge private) at endline. Means are calculated within cells formed by grade group (pre-primary, lower primary, and upper primary), subsample (*main*,*brig*,*priv*), and bridge location. To obtain final predictions, we apply empirical Bayes shrinkage to the cell means.[65]

The results of the overidentified 2SLS specification suggest that there is no negative effect of paying fees on test scores, and if paying fees has any effect on test scores, it is positive. At the primary school level, the effect of Bridge enrollment is between 0.95 and 1.06 EYS, depending on the instrument set, and the effect of paying fees is between 0.03 and 0.04 EYS per KES 1,000 (Panel A, Table 13). At the pre-primary level, the effect of Bridge enrollment is between 1.90 and 2.01 EYS and the effect of paying fees is between 0.07 and 0.08 EYS per KES 1,000.

The positive effect is surprising and warrants some discussion. There are at least two reasons why the effect could be positive: schools may focus resources on paying customers; and pupils may feel obliged to study harder; parents may put more pressure on their children to study when they are paying fees. The positive effect could also reflect a violation of the LATE heterogeneity with fees paid in the absence of the scholarship homogeneity assumption required for identification of the causal effect (Hull, 2018). If the LATE is larger for applicants who are more or less likely to pay high fees in the absence of the scholarship, the 2SLS estimates of the effect of paying fees will reflect that heterogeneity.

8 Effectiveness of non-Bridge private schools

A 2SLS model with two endogenous variables representing non-Bridge private school enrollment and public school enrollment suggests that Bridge improves test scores relative to other private schools. These results are consistent with other work in Kenya suggesting that pupils who enroll in low-cost private schools do not gain more on test scores than those in public schools in Kenya (Zuilkowski et al., 2020). As in Section 7.2, the model is identified with multiple instruments formed by interacting the scholarship offer with baseline characteristics. Under the assumption that the baseline characteristics used to form the interactions are not predictive of heterogeneity

[65] The empirical Bayes shrinkage estimator models the binary choice to enroll in Bridge in 2017 using a Beta prior, a conjugate prior for a binomial posterior. See Gray-Lobe et al. (2020) for details on the procedure.

in the relative effectiveness of each sector (Hull, 2018). The separate effects of public schools and other private schools compared to Bridge are identified by variation in the first stage effect of the scholarship on enrollment in each sector across applicants with different baseline characteristics.

At the primary school level, we pool the P_{main} and the P_{priv} samples because the combined sample has more variation in the effects of the scholarship on public and non-Bridge private school enrollment. At the pre-primary level, we focus on the PP_{main} sample because this sample already contains pupils who were planning to enroll at private schools.[66]

At the primary school level, our preferred specification – that for which F-statistics suggest that bias due to weak instruments may be small – (Panel B of Table 13) supports the view that Bridge is more effective than other private schools. Using the academy-intention predictions to form the instruments (Column 2), the estimated effect of Bridge relative to other private schools relative to Bridge is 0.89 EYS. The multivariate F-statistic is 11.39 and the Angrist-Pischke partial F-statistic for the Bridge vs. other private schools is 16.67. One caveat is that this is only marginally statistically significant. Another caveat is that in another specification, using the intention and demographic interactions as instruments (Column 3), the estimated effect of enrolling in a non-Bridge private school compared to Bridge is approximately zero and imprecisely estimated. However, in this alternative specification, the multivariate F-statistic is only 4.82 and the Angrist-Pischke partial F-statistic for the Bridge vs. other private schools effect is only 5.80, raising concerns of bias due to weak instruments.

At the pre-primary level, all specifications find positive and statistically significant effects of enrolling at Bridge schools relative to other private schools. The estimates range from 0.23 to 1.80 EYS. While this pattern is consistent with that in primary school, we note that the estimates are imprecise and the F-statistic is below 2 for all specifications, raising weak instrument concerns.

While our preferred estimates of EYS gains suggest that students who move from other private schools to Bridge score more highly on tests of subject matter knowledge, use of standardization would seem worth exploring in many situations even if learning gains were equivalent in Bridge and in "mom and pop" private schools at the time of the study. For example, while managers of "mom and pop" private schools typically own the school and therefore face high-powered incentives,

[66] We do not include the samples with pupils who were already in Bridge because these samples complicate the interpretation of the endogenous variable insofar as the scholarship reduced dropout over time as opposed to encouraging enrollment in the first year of the program.

managers of public schools where education is free are not residual claimants with high-powered incentives and standardization could help substitute for these high-powered incentives. Another scenario in which a standardized system can be useful is in the context of environments, where subnational or national graduates of teacher training colleges are scarce, and secondary school graduates already teach, particularly at the pre-primary level. A standardized system which helps secondary school graduates function effectively may be useful in these environments. It is also worth bearing in mind that there may be significant scope for standardized models to be refined and improved over time through A/B testing, for example. At the time of our study, Bridge was new and had not had much opportunity to undertake such processes. Standardized models may thus be worth exploring even if pupils performed equally well in Bridge and in other private schools at the time of our study.

9 The Movement from Standardization within Private Schools to Standardization in Public Schools

Since the time of the study, Bridge has closed 293 of its 405 private schools in Kenya.[67] Bridge's parent company, NewGlobe, made an initial foray into working with governments in a pilot of a charter-school like model in Liberia, but then moved the great majority of its work to a model in which it acts as a service provider for governments.Over 95 percent of pupils served by NewGlobe are now reached through public schools (NewGlobe, email communication, June 2, 2022).

Under the Bridge private school model, standardization was used with teachers who had less education and experience than public school teachers and received much lower compensation. In the new model, standardization is used with public school teachers with unchanged qualifications and compensation, much as checklists are used with fully-trained pilots and surgeons. Moreover, under this approach the state, rather than parents, pay for the services of NewGlobe.

From a positive social science standpoint, it is worth considering three possibilities for why most Bridge private schools in Kenya closed, despite the test score gains they generated. First, there might have simply been insufficient demand for Bridge schools independent of the policy

[67]The normative question of whether the closure of most Bridge private schools was desirable depends on a wide range of factors that are beyond the scope of this paper, including judgements on various concerns raised about Bridge, which cannot be adjudicated with the data we analyze.

environment. Parental demand for school may be influenced by many factors beyond impacts on test scores, and perhaps only a limited number of parents want the Bridge model, with its use of less educated teachers, long hours, and very basic physical structures. As discovered below, many concerns have been raised about Bride, and many parents may have similar concerns. In our main sample almost half of the lottery winners at the primary school level turned down the scholarship in 2017. Thus it is possible that independent of the policy environment, Bridge would only have enough consumers to keep a limited number of schools open. However, the environment faced by Bridge private schools also changed. This can be understood in the context of two broader perspectives from social science. First, Hsieh and Klenow (2009) argue that regulatory, tax, and labor market institutions create "wedges" in low- and middle-income countries that constrain the growth of high productivity private firms because small "mom and pop" private firms can operate under the radar screen and are thus less subject to these wedges while governments effectively subsidize state-owned firms. A second perspective also sees social choices around the adoption of technology as important, but views this in a more positive light. From this perspective, there is a historical pattern in which standardization of production has often led to social conflict, particularly when it involves hiring workers with less education and experience and lower compensation. This social conflict sometimes leads to the formation of a new social contract, in which standardization is adopted under institutions chosen by society to advance social values. For example, mass production in the automobile industry replaced artisanal production but eventually led to the rise of the United Auto Workers (UAW), a wave of strikes and lockouts, and the 1935 National Labor Relations Act which institutionalized a new social contract under which much of U.S. manufacturing would be subject to collective bargaining. Consistent with the latter two perspectives, civil society and labor groups shared concerns about Bridge, and policymakers subsequently changed their policy. This can arguably be seen as a successful example of a broader recent labor movement strategy, sometimes termed "alt-labor", in which rather than relying exclusively on organizing workers to join unions and then engaging in collective bargaining, the labor movement also seeks to work with customers, investors, and regulators to influence firms (Walsh, 2018).

The history of World Bank's involvement with Bridge helps illustrate how engagement by civil society and labor organizations may have shaped the environment faced by Bridge and NewGlobe. Prior to the period of our study, the World Bank's International Finance Corporation (IFC) had

invested directly in Bridge (IFC, 2014), and Jim Yong Kim, the President of the World Bank praised Bridge in a 2015 speech (World Bank, 2015). Subsequent to Jim Yong Kim's speech, 116 civil society organizations signed a joint letter calling for the World Bank to "stop promoting the model used by Bridge International Academies and other fee-charging, private schools, and publicly re-commit the World Bank to universal, free and compulsory basic education" (Education International, 2015).

In 2018, the East Africa Centre for Human Rights (EACHRights), a Kenyan NGO, lodged a formal complaint against the IFC with the Office of the Compliance Advisor Ombudsman (CAO) alleging that Bridge violated IFC's Performance Standards as well as national and international law. This complaint included several allegations against Bridge's operational model in Kenya (CAO, 2018).[68]. [69] EACHRights later supported parents in reporting to the CAO that two Bridge pupils were electrocuted (in one case fatally) after touching an electric wire hanging in a building adjacent to the school compound (CAO, 2020a,b). There were also allegations of child sexual abuse involving Bridge students and staff (CAO, 2020c). [70] These complaints are in different phases of dispute resolution or compliance at the office of CAO.

While the most recent complaints remain under investigation by the CAO, the policy of World Bank/IFC changed through another avenue. In 2020, the World Bank sought a replenishment for the International Development Association (IDA), which provides concessional development financing to many lower income countries. This required U.S. congressional approval.

Labor and Civil society groups reached out to U.S. officials. Education International (EI), a body bringing together organizations of teachers internationally, refers on its website to "intense lobbying with Congress and successive Administrations by the AFT and the NEA; firm, sustained support of EI officers and the Executive Board and the ITUC/Global Unions office in Washington, DC." (Education International, 2020). The Chair of the House Financial Services Committee (House Financial Services Committee, 2020) advocated a package which included a commitment

[68]The allegations included 'economic discrimination, discrimination of students with disability, poor quality and lack of compliance with the national curriculum, violation of the rule of law, violation of health and safety regulations, intimidation of stakeholders, lack of transparency, lack of parental inclusion, advertising of false or misleading information, excessive working hours, payment below minimum wage, and other labour issues' prevalent in Bridge schools.

[69]The IFC responded to this complaint by stating that it "stands by its investment in Bridge but that it will support the CAO investigation and 'take corrective action as needed"' (Devex, 2019).

[70]For a report commissioned by Bridge on the issue of child sexual abuse, see Tunza Child Safeguarding (2020).

to "an official freeze of any direct or indirect investments in private for-profit K-12 schools". As of 2022, the IFC has 'exited its direct investment' in Bridge, as part of a broader decision to stop funding for-private K-12 schools (IFC, 2022). The World Bank thus changed its policy not just towards Bridge private schools but to private K-12 schools more generally.

The Kenyan National Union of Teachers (KNUT) also communicated concerns about Bridge's private school model widely in Kenya, including in the media, potentially influencing both, potential Bridge customers and government regulators.

Media statements by the leader of the KNUT and regulatory uncertainty surrounding Bridge, covered in the Kenyan media (Herbling, 2015; Hengeveld, 2017; The Standard, 2017), may have reduced parents' willingness to enroll their children at Bridge. Some officials at Kenya's Ministry of Education Science and Technology argued that Bridge needed to adhere to private school regulations outlined in the 2013 Education Act, which mandates outdoor playgrounds with provisions for children with disabilities; specific types of toilet, kitchen and dining facilities; and standard size classrooms. Bridge argued that it is a 'non-formal' school and thus comes under the purview of the Alternative Provision of Basic Education and Training (APBET) guidelines (the regulations for which are not as stringent as they are for private schools) and has substantially complied with its registration requirements (Bridge, 2016). Others argued that Bridge's curriculum did not conform to standards set by the Kenya Institute of Curriculum Development (KICD), while Bridge argues that they conform to the Kenyan curriculum and their materials have been conditionally approved by KICD, awaiting final approval (Bridge, 2016) (Our study finds that Bridge pupils performed well on tests designed around the Kenyan curriculum, but we did not assess whether the material in Bridge lesson plans corresponds to that in the national curriculum). Kenyan governmental regulations required Bridge to hire more certified teachers over time and although Bridge pay scales do not depend on certification, the supply of certified teachers is smaller than that of secondary school graduates so this likely limited Bridge's flexibility in making human resource decisions.

As discussed earlier, this study can only provide limited evidence on these controversies, specifically around teacher pay and conditions,safety of the school grounds, and the effect of school fees on missed classes, and it is beyond the scope of this paper to adjudicate disputes between Bridge and its critics or take a stand on the appropriateness of policy decisions by international organizations, and national regulators regarding Bridge private schools.

At this point, we also do not have sufficient evidence to distinguish between the Hsieh and Klenow (2009) view, in which wedges limiting firm growth prevent productivity gains that large firms could create, destroying surplus, and the more positive view that political action can lead to outcomes in which technology is used in a way consistent with social preferences, but with a socially chosen division of a similar surplus. Distinguishing between the two views will depend in part on assessing the gains from standardization in public school systems.

From a policy perspective, a critical next step for research is to assess the impact of standardization within public schools, since at least for the time being, it seems unlikely that a substantial share of children will be educated through private schools employing the approach to standardization used by Bridge, while NewGlobe is instead rapidly expanding the number of children it reaches through provision of services to the public sector.

10 Conclusion

We use random variation in scholarship assignment from an oversubscribed program to evaluate the effect of enrolling in schools that employ a highly-structured and standardized approach to education. This included delivering detailed lesson plans to teachers on tablet computers; providing feedback and monitoring to teachers and establishing detailed standardized procedures for a host of other activities, from school construction to handling finances.

Enrolling in these schools speeds grade progression, increases scores on a test of academic knowledge across a range of subjects at a range of grade levels, reduces dispersion in scores, and improves an array of non-curricular cognitive outcomes such as working memory, self-regulation and receptive vocabulary.

Test score gains are similar across locations, consistent with the hypothesis that the standardized operations lead to less dispersed outcomes. The financial transfer to households that would have attended private schools in any case is unlikely to be the main driver of the results.

The study provides some limited evidence on areas where concerns have been raised about the provider, Bridge International Academies, including teacher qualifications, HR practices, child safety, and the impact of school fees on access to education. Less than one-fourth of the teachers employed by Bridge had more than secondary school education compared to three-quarters of

teachers in public and private schools. Relative to public school teachers, Bridge teachers were younger, less experienced, and worked longer hours, including Saturdays. School staff received compensation that was a fraction of that of public school teachers, and about the same as that of teachers in private schools attended by the population we study. The study found that primary school pupils who enrolled at Bridge were eight percentage points more likely to report that their playing fields have hazards (significant at the 10 percent level). They were six percentage points less likely to report that their teachers engaged in corporal punishment (although this difference is only marginally significant). At the primary level, among scholarship non-recipients, Bridge pupils were more likely to report missing school due to inability to pay fees than public school pupils.

One caveat is that we examine a population of pupils who applied for a scholarship for these schools. Effects could differ for other populations. Another caveat is that important changes have taken place since the time of this study. Kenya has adopted a new Competency-Based Curriculum, which features more frequent "continuous" standardized assessments (Akala, 2021), supported by a new data management system (Gitonga, 2018). Roughly contemporaneous with this study, the government also launched a structured pedagogy to help improve early grade learning in foundational subjects[71]. Bridge's parent company, New Globe, shut down most of its private schools in Kenya and now overwhelmingly works as a service provider to governments operating public schools.

Beyond the specific case of Bridge, this study shows that attending schools delivering highly standardized education has the potential to produce dramatic learning gains at scale, suggesting that policymakers may wish to explore incorporation of standardization, including standardized lesson plans and teacher feedback and monitoring, in their own systems.

Since the effects of standardization in education could vary between public schools and Bridge private schools, with their very different human resource policies, as well as with local circumstances and implementation, further research on efforts to introduce more structure into public education systems would be needed to understand the impact of introducing a high degree of structure in education in other contexts.

[71]In 2015, the Kenya Tusome Early Grading Activity was launched nationally to support early grade learning among Grade 1 students, which expanded to Grade 3 students in 2017 (Piper et al., 2018)

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

Abadie, A., J. D. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70.

Abadie, A., M. M. Chingos, and M. R. West (2018): "Endogenous Stratification in Randomized Experiments," *The Review of Economics and Statistics*, 100.

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011): "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots," *The Quarterly Journal of Economics*.

Akala, B. M. (2021): "Revisiting education reform in Kenya: a case of competency based curriculum (CBC)," *Social Sciences & Humanities Open*, 3, 100107.

Anderson, J. (2018): "The controversial Silicon Valley-funded quest to educate the world's poorest kids," *Quartz January*, 22, 2018.

Andrews, I., J. Stock, and L. Sun (2019): "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics*.

Angrist, J. D., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002): "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment," *American Economic Review*, 92, 1535–1558.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton.

Angrist, N., D. K. Evans, D. Filmer, R. Glennerster, H. F. Rogers, and S. Sabarwal (2020): "How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric," Working paper, World Bank.

Arrow, K. J. (1974): *The limits of organization*, WW Norton & Company.

Autor, D. (2014): "Polanyi's Paradox and the Shape of Employment Growth," Working Paper 20485, National Bureau of Economic Research.

Baker, M., J. Gruber, and K. Milligan (2008): "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *Journal of Political Economy*, 116, 709–745.

Banerjee, A., S. Cole, E. Duflo, and L. Linden (2007): "Remedying Education: Evidence from Two Randomized Expeirments in India," *The Quarterly Journal of Economics*, 122, Link.

Banerjee, A. and E. Duflo (2014): "Economics education in the digital age: The implications of online technologies and MOOCS," in *American Economic Review: Papers and Proceedings*, Citeseer, vol. 104, 514–518.

Banerjee, A., S. Jacob, M. Kremer, J. Lanjouw, and P. Lanjouw (2005): "Moving to Universal Education: Costs and Tradeoffs," *Mimeogr., Department of Economics, MIT*.

Bau, N. and J. Das (2021): "Teacher Value Added in a Low-Income Country," *American Economic Journal: Economic Policy*, 12, 1.

Beck, S. and S. May (2016): "*Low-cost private schools scalable solution to a massive gap?*" Link [Accessed: May 20, 2020].

Behrman, J. R. and S. W. Parker (2013): "Is Health of the Aging Improved by Conditional Cash Transfer Programs? Evidence from Mexico," *Demography*, 50, Link.

Berkes, J. and A. Bouguen (2022): "Estimating Preschool Impacts with close substitute : Bounds, conditional LATE and Predicted sub-LATE," .

Berry, J., H. Kannan, S. Mukherji, and M. Shotland (2020): "Failure of frequent assessment: An evaluation of India's continuous and comprehensive evaluation program," *Journal of Development Economics*, 143, 102406.

Bertrand, M. and A. Schoar (2006): "The role of family in family firms," *Journal of economic perspectives*, 20, 73–96.

Blimpo, M. P., P. Carneiro, P. Jervis, and T. Pugatch (2018): "Improving Access and Quality in Early Childhood Development Programs: Experimental Evidence from the Gambia," *IZA Discussion Paper Series*.

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

Bold, T., D. Filmer, G. Martin, E. Molina, B. Stacy, C. Rockmore, J. Svensson, and W. Wane (2017): "Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa," *Journal of Economic Perspectives*, 31, 185–204.

Bold, T., D. Filmer, E. Molina, and J. Svensson (2019): "The Lost Human Capital: Teacher Knowledge and Student Achievement in Africa," *Working Paper*.

Bold, T., M. Kimenyi, G. Mwabu, and J. Sandefur (2011): "The High Return to Private Schooling in a Low-Income Country," Tech. rep., Brookings Institution, Africa Growth Initiative Working Paper, Washington, DC, Link.

——— (2014): "Can Free Provision Reduce Demand for Public Services? Evidence from Kenyan Education," *World Bank Economic Review*, 29, 293–326, Link.

Bridge (2016): "Response to Bridge vs Reality: A study of Bridge International Academies' for profit schooling in Kenya," Bridge International Academies, Link.

——— (2017): "Bridge's Open Letter Response to Those Calling to Stop Investment in Proven Solutions," Bridge International Academies, Link.

CAO (2018): "Kenya / Bridge International Academies-01/Kenya," Compliance Advisor Ombudsman, IFC, Link.

——— (2020a): "Kenya / Bridge International Academies-02/Kenya," Compliance Advisor Ombudsman, IFC, Link.

——— (2020b): "Kenya / Bridge International Academies-03/Kenya," Compliance Advisor Ombudsman, IFC, Link.

——— (2020c): "Kenya / Bridge International Academies-04/Vice President request," Compliance Advisor Ombudsman, IFC, Link.

Chernozhukov, V. and C. Hansen (2013): "Quantile Models with Endogeneity," *Annual Review of Economics*.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104, 2633–79.

Choi, H.-Y., E. H. Kim, J. Yoo, K. Lee, D. Song, young Dae Kim, H.-J. Cho, H. S. Nam, K. Y. Lee, H. S. Lee, and J. H. Heo (2016): "Decision-Making Support Using a Standardized Script and Visual Decision Aid to Reduce Door-to-Needle Time in Stroke," *Journal of Stroke*, 18, 239–241.

Cohodes, S., E. Setren, and C. R. Walters (2019): "Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector," Working Paper 25796, National Bureau of Economic Research.

Cristia, J., P. Ibarrarán, S. Cueto, A. Santiago, and E. Severín (2017): "Technology and child development: Evidence from the one laptop per child program," *American Economic Journal: Applied Economics*, 9, 295–320.

Currie, J. (2001): "Early Childhood Education Programs," *Journal of Economic Perspectives*, 15(2), 213–238.

Dean, E. B., F. Schilbach, and H. Schofield (2018): "Poverty and Cognitive Function," in *The Economics of Poverty Traps*, ed. by C. B. Barret, M. R. Carter, and J.-P. Chavas, University of Chicago Press, 57–118, Link.

Dean, J. T. and S. Jayachandran (2019): "Attending kindergarten improves cognitive but not socioemotional development in India," Unpublished.

Devercelli, A. E. and R. K. Sayre (2016): "Scaling up preschool in Kenya: costs, constraints, and opportunities," Tech. rep., The World Bank.

Devex (2019): "CDC supports expansion of Bridge International Academies with US$6 million investment," .

DHS (2014): "Kenya Demographic and Health Survey," Tech. rep., Kenya National Bureau of Statistics, Nairobi, Kenya, Link.

Dresser, R. (2012): "The Impact of Scripted Literacy Instruction on Teachers and Students," *Issues in Teacher Education*.

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

——— (2015b): "School governance, teacher incentives, and pupil-teacher ratios: experimental evidence from Kenyan primary schools," *Journal of Public Economics*, 123, 92–110.

——— (2021): "The Impact of Free Secondary Education: Experimental Evidence from Ghana," Working Paper 28937, National Bureau of Economic Research.

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 102, 1241–1278.

Eble, A., C. Frost, A. Camara, B. Bouy, M. Bah, M. Sivaraman, P.-T. J. Hsieh, C. Jayanty, T. Brady, P. Gawron, et al. (2021): "How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in The Gambia," *Journal of Development Economics*, 148, 102539.

Education International (2015): ""Just" $6 a month?: The World Bank will not end poverty by promoting fee-charging, for-profit schools in Kenya and Uganda," Link.

——— (2020): ""World Bank shifts position on funding for most private-for-profit schools"," .

Education International and KNUT (2016): "Bridge vs. Reality: A study of Bridge International Academies' for profit schooling in Kenya," Education International and Kenya National Union of Teachers.

Eifert, G. H., D. Schulte, M. J. Zvolensky, C. Lejuez, and A. W. Lau (1997): "Manualized behavior therapy: Merits and challenges," *Behavior Therapy*, 28, 499–509.

Evans, D. and A. Popova (2016): "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews," *World Bank Research Observer*, 31.

Evans, D. and F. Yuan (2019): "Equivalent years of schooling: A metric to communicate learning gains in concrete terms," *World Bank Policy Research Working Paper*.

——— (2020): "How big are effect sizes in international education studies," *Center for Global Development, Working Paper*, 545.

Evans, D. K. and M. Ngatia (2020): "School Uniforms, Short-Run Participation, and Long-Run Outcomes: Evidence from Kenya," *The World Bank Economic Review*.

Fazzio, I., A. Eble, R. L. Lumsdaine, P. Boone, B. Bouy, P.-T. J. Hsieh, C. Jayanty, S. Johnson, and A. F. Silva (2021): "Large learning gains in pockets of extreme poverty: Experimental evidence from Guinea Bissau," *Journal of Public Economics*, 199, 104385.

Friedman, W., M. Kremer, E. Miguel, and R. Thornton (2016): "Education as liberation?" *Economica*, 83, 1–30.

Gitonga, A. (2018): "Education: State launches digital registration of learners ," Link.

Goodenough, F. (1926): *Measurement of intelligence by drawings*, New York, New York: World Book Co.

Gray-Lobe, G., A. Keats, M. Kremer, I. Mbiti, and O. Ozier (2020): *Evaluation of Bridge International Academies in Kenya: Stage One Analysis and Plan for Subsequent Analysis*, AEA RCT Registry, Link.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2021): "The Long-Term Effects of Universal Preschool in Boston," Working Paper 28756, National Bureau of Economic Research.

Guilford, J. P. (1967): *The Nature of Human Intelligence*, New York, New York: McGraw-Hill.

Harris, D. B. (1963): *Children's drawings as measures of intellectual maturity*, New York, New York: Harcourt, Brace, & World, Inc.

Haushofer, J. and J. Shapiro (2016): "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *The Quarterly Journal of Economics*, Link.

Haynes, A. B., T. G. Weiser, W. R. Berry, S. R. Lipsitz, A.-H. S. Breizat, E. P. Dellinger, T. Herbosa, S. Joseph, P. L. Kibatala, M. C. M. Lapitan, A. F. Merry, K. Moorthy, R. K. Reznick, B. Taylor, and A. A. Gawande (2009): "A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population," *New England Journal of Medicine*, 360, 491–499, pMID: 19144931.

Heckman, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limitd Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475–492.

Hengeveld, M. (2017): "No education crisis wasted: On Bridge's "business model" in Africa (Translation)," *De Correspondent/ Africa Is A Country*.

Herbling, D. (2015): "Ministry Stops Expansion of Low-Cost Schools," *Business Africa Daily*, Link.

House Financial Services Committee (2020): "Chairwoman Waters Applauds World Bank Capital Increase Included In COVID-19 Stimulus Package," .

Hoxby, C. M. and S. Murarka (2009): "Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement," Working Paper 14852, National Bureau of Economic Research.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124.

Hsieh, C.-T. and B. A. Olken (2014): "The missing" missing middle"," *Journal of Economic Perspectives*, 28, 89–108.

Hsieh, C.-T. and M. Urquiola (2006): "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of Public Economics*, 90, 1477–1503.

Huillery, E., J. de Laat, and P. Gertler (2017): "Supporting Disadvantaged Children to Enter Kindergarten: Experimental Evidence from Bulgaria," *World Bank Impact Evaluation Report*, 116.

Hull, P. (2018): "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons," Manuscript.

Humphreys, M., R. S. De la Sierra, and P. Van der Windt (2013): "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration," *Political Analysis*, 21, 1–20.

IFC (2014): "IFC Supports Expansion of Primary Education with Investment in Bridge International Academies," .

——— (2022): "Bridge International Academies," Project Disclosures, IFC, Link.

Ilias, N. (2006): "Families and firms: Agency costs and labor market imperfections in Sialkot's surgical industry," *Journal of Development Economics*, 80, 329–349.

Imbens, G. W. and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and C. F. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Ingersoll, R. M. and T. M. Smith (2003): "The wrong solution to the teacher shortage," *Educational leadership*, 60, 30–33.

Jackson, C. K. (2018): "What do test scores miss? The importance of teacher effects on non–test score outcomes," *Journal of Political Economy*, 126, 2072–2107.

Johnson, S. M., M. A. Kraft, and J. P. Papay (2012): "How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement," *Teachers college record*, 114, 1–39.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *The Quarterly Journal of Economics*, 131.

Kline, P. and C. R. Walters (2016): "Evaluating Public Programs with Close Substitutes: the Case of Head Start," *The Quarterly Journal of Economics*, 141, 1795–1848.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75, 83–119.

Knauer, H. A., P. Kariger, P. Jakiela, O. Ozier, and L. C. Fernald (2019): "Multilingual assessment of early child development: Analyses from repeated observations of children in Kenya," *Developmental Science*, 22, e12875.

Knutsson, D. and B. Tyrefors (2022): "Quality and Efficiency Between Public and Private Firms: Evidence From Ambulance Services," *The Quarterly Journal of Economics*.

Kolen, M. J. and R. L. Brennan (2014): *Test Equating, Scaling, and Linking*, New York, New York: Springer-Verlag.

Kremer, M., C. Brannen, and R. Glennerster (2013): "The Challenge of Education and Learning in the Developing World," *Science*, 340.

Kremer, M. and A. Holla (2009): "Improving education in the developing world: what have we learned from randomized evaluations?" *Annu. Rev. Econ.*, 1, 513–542.

Kremer, M., G. Rao, and F. Schilbach (2019): "Behavioral Development Economics," in *Handbook of Behavioral Economics - Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier B.V.

Kwauk, C. and J. P. Robinson (2016): "Bridge International Academies: Delivering Quality Education at a Low Cost in Kenya, Nigeria, and Uganda," .

Lamb-Sinclair, A. (2017): "Why Teachers Need Their Freedom," *The Atlantic*.

Leaver, C., O. Ozier, P. Serneels, and A. Zeitlin (2021): "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools," *American Economic Review*, 111, 2213–46.

Lemos, R., K. Muralidharan, and D. Scur (2021): "Personnel management and school productivity: Evidence from india," *National Bureau of Economic Research Working Paper*.

Lipsey, M. W., D. C. Farran, and K. Durkin (2018): "Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade," *Early Childhood Research Quarterly*, 45, 155–176.

Lucas, A. M. and I. M. Mbiti (2012): "Access, Sorting, and Achievement: The short-run effects of Free Primary Education in Kenya," *American Economic Journal: Applied Economics*, 4.

Marshall, A. (1919): *Industry and Trade*, London.

Martin, G. H. and O. Pimhidzai (2013): "Service Delivery Indicators: Kenya," Tech. rep., The World Bank, Washington, DC, Link.

McClelland, M. M., C. M. Connor, A. M. Jewkes, C. E. Cameron, C. L. Farris, and F. J. Morrison (2007): "Links Between Behavioral Regulation and Preschoolers' Literacy, Vocabulary, and Math Skills," *Developmental Psychology*, 43, 947–959, Link.

Ministry of Education (2016a): "2016 Basic Education statistical Booklet," Link.

——— (2016b): "MOE.HQS/3/6/56 RE: Registration of Bridge International Academies (BIA) LTD," Link.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019): "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 109, 1426–60.

Muralidharan, K. and V. Sundararaman (2013): "Contract teachers: Experimental evidence from India," Tech. rep., National Bureau of Economic Research.

——— (2015): "The Aggregate Effect of School Choice: Evidence from a Two-stage Experiment in India," *The Quarterly Journal of Economics*, 130, 1011–1066.

Nation (2020): "TSC ordered to register all ECDE teachers recruited by counties," Link.

Ngware, M., H. P. P. Donfouet, N. Hungi, and P. Kitao-Wekulo (2018): "Cost and Cost-Effectiveness of a Pre-Primary School Program in Kenya," *Economics Working Paper Archive (University of Rennes 1 and University of Caen)*, Link.

Nunnery, A. (2021): "Teacher Autonomy Matters. Here's Why." *Medium*.

Olken, B. A. (2015): "Promises and Perils of Pre-analysis Plans," *Journal of Economic Perspectives*, 29.

Ozier, O. (2018a): "Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming," *American Economic Journal: Applied Economics*, 10, 235–62.

——— (2018b): "The impact of secondary schooling in Kenya: a regression discontinuity analysis," *Journal of Human Resources*, 53, 157–188.

Piper, B., S. S. Zuilkowski, D. Kwayumba, and A. Oyanga (2018): "Examining the secondary effects of mother-tongue literacy instruciton in Kenya: Impacts on student learning in English, Kiswahili, and mathematics," *International Journal of Educational Development*, Link.

Piper, B., S. S. Zuilkowski, and A. Mugenda (2014): "Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative," *International Journal of Educational Development*.

Plucker, J. A. (1999): "Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) Longitudinal data," *Creativity Research Journal*, 12.

Polanyi, M. (1967): *The Tacit Knowledge Dimension*, London: Routledge.

Rangan, V. K. and K. Lee (2010): "Bridge International Academies: A School in a Box," *Harvard Business School Case Study*.

Reardon, S. F. and S. W. Raudenbush (2013): "Under What Assumptions Do Site-by-Treatment Instruments Identify Average Causal Effects," *Sociological Methods & Research*, 42, 143–163.

Riep, C. and M. Machacek (2016): "Schooling the Poor Profitably: the Innovations and Deprivations of Bridge International Academies in Uganda," Education International.

Romero, M. and J. Sandefur (2021): "Beyond Short-Term Learning Gains: The Impact of Outsourcing Schools in Liberia After Three Years*," *The Economic Journal*, ueab087.

Romero, M., J. Sandefur, and W. A. Sandholtz (2020): "Outsourcing Education: Experimental Evidence from Liberia," *American Economic Review*, 110, 364–400.

Sandefur, J. (2016): "Internationally Comparable Mathematics Scores for Fourteen African Countries," Working Paper 444, Center for Global Development.

Stockard, J., T. W. Wood, C. Coughlin, and C. R. Khoury (2018): "The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research," *Review of Educational Research*, 88, 479–507.

The National Treasury (2018): "Education Sector Report," Link.

The Standard (2017): "Storm over plan to close down 12 private schools in Busia," Link.

Torrance, E. P. (1972): "Predictive validity of the Torrance Tests of Creative Thinking," *Journal of Creative Behavior*, 6.

TSC-KNUT CBA (2016): "Certificate of Registration of Collective Bargaining Agreement Between Teachers Service Commission and Kenya National Union of Teachers (KNUT)," Link.

Tunza Child Safeguarding (2020): "Keeping Pupils Safe in Kenya: A Review and Assessment of Child Safeguarding in Kenya," Link.

Uwezo (2016): "Are Our Children Learning? Uwezo Kenya Sixth Learning Assessment Report," Tech. rep., Nairobi: Twaweza East Africa, Nairobi, Kenya, Link.

Valencia, S. W., N. A. Place, S. D. Martin, and P. L. Grossman (2006): "Curriculum Materials for Elementary Reading: Shackles and Scaffolds for Four Beginning Teachers," *The Elementary School Journal*, 107.

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

Walsh, D. (2018): "Alt-labor, explained," *MIT Sloan School of Management – Ideas Made to Matter*, Link.

Walters, C. R. (2015): "Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start," *American Economic Journal: Applied Economics*, 7, 76–102.

Weiland, C. and H. Yoshikawa (2013): "Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills," *Child Development*, 84, 2112–2130.

Wilson, G. (1996): "Manual-based treatments: The clinical application of research findings," *Behaviour Research and Therapy*, 34, 295–314.

World Bank (2015): "Speech by World Bank Group President Jim Yong Kim: Ending Extreme Poverty by 2030: The Final Push," Link.

——— (2020): "The Human Capital Project," .

Wüthrich, K. (2020): "A Comparison of Two Quantile Models with Endogeneity," *Journal of Business & Economic Statistics*, 38, 443–456, Link.

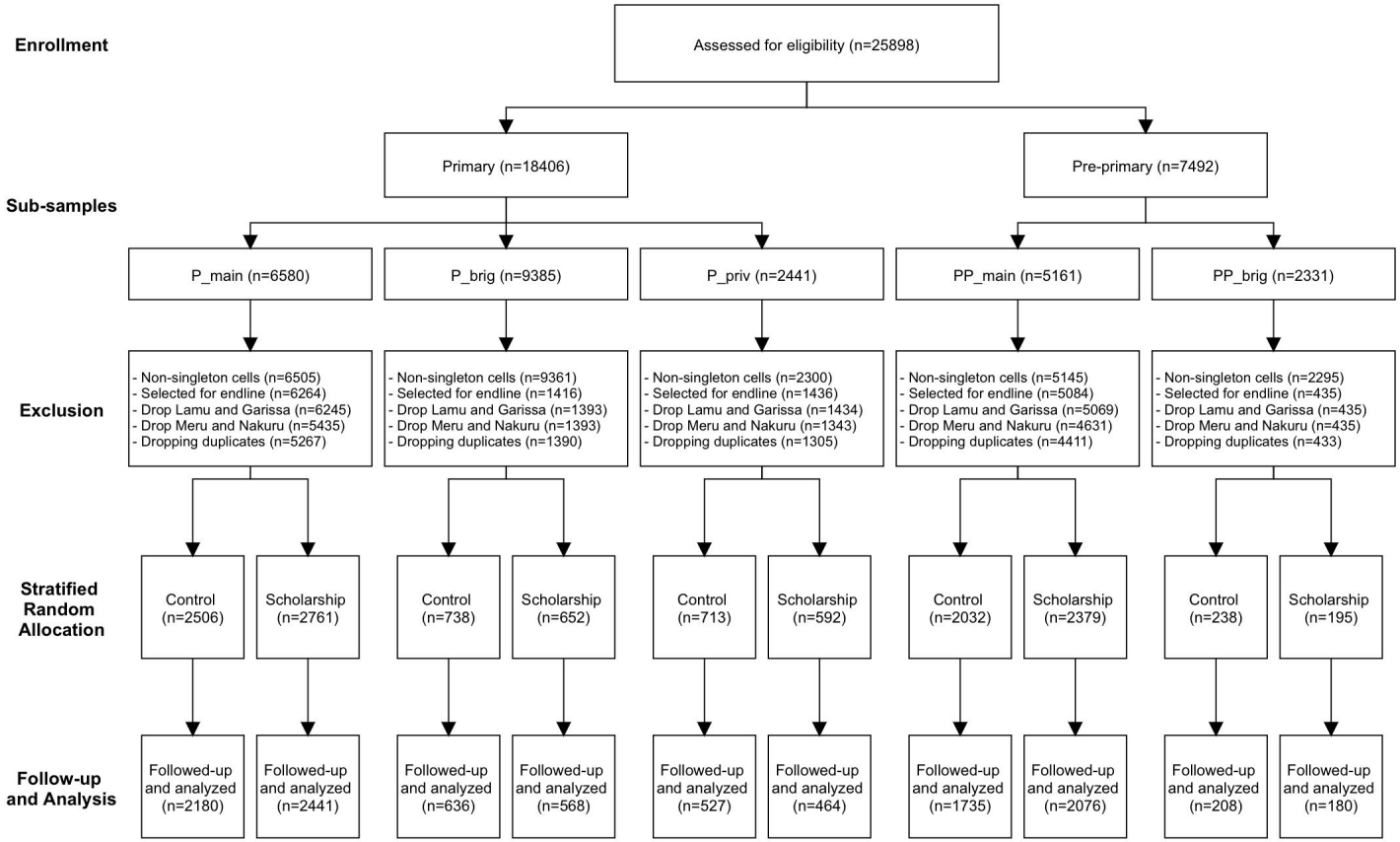
Yando, R., V. Seitz, and E. Zigler (1979): *Intellectual and personality characteristics of children: Social-class and ethnic-group differences*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Zuilkowski, S. S., B. Piper, and S. O. . O. Kiminza (2017): "Parents, quality, and school choice: why parents in Nairobi choose low-cost private schools over public schools in Kenya's free primary education era," *Oxford Review of Education*.

Zuilkowski, S. S., B. Piper, and S. Ong'ele (2020): "Are Low-Cost Private Schools Worth the Investment?: Evidence on Literacy and Mathematics Gains in Nairobi Primary Schools." *Teachers College Record*, 122, n1.

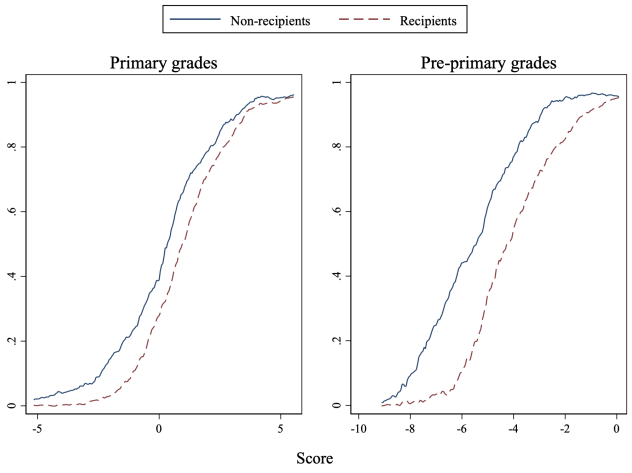
Figures

Figure 5: Study design



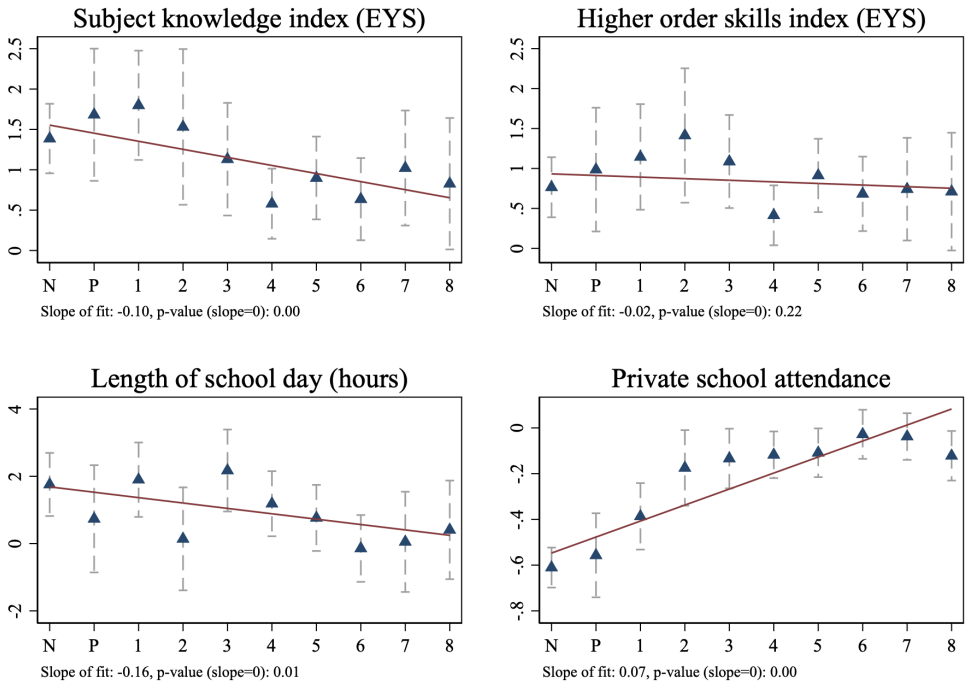
Notes: The figure shows how the final sample was constructed. The first row shows the total number of applications received that were included in the randomized scholarship lottery after initial restrictions including removing twenty duplicates, two incomplete applications, 515 applications for pupils who received scholarships through other programs, and applying eligibility critieria described in the text. The second row restricts to randomization strata that included more than one applicant. *Non-singleton cells* refer to randomization strata that included more than one applicant. *Selected for endline* refers to removing randomly selected applications as described in Section 3.4. Applicants from Lamu and Garissa were not included in endline activities due to security constraints. Applications received through UWR in the counties of Meru and Nakuru were removed due to quality concerns. Duplicate applications for the same child were identified during endline and dropped.

Figure 6: Estimated distributions of aggregate subject knowledge index



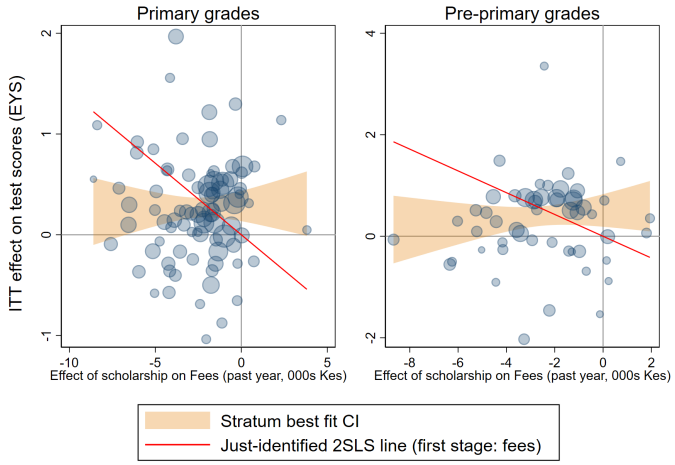
Notes: Figure shows the estimated distribution of test scores for scholarship recipients and non-recipients for the subset of compliers using an approach from Abadie (2002). Each point in the estimated CDF reflects a separate 2SLS estimate. The recipient CDF gives the 2SLS estimated effect of enrolling at Bridge $D_i \in \{0, 1\}$ on the transformed dependent variable $W_i^D = \mathbb{1}\{Y_i > x\}D_i$ where x represents a point in the support of the test score distribution. The non-recipient CDF gives the 2SLS estimated effect of not enrolling at Bridge $1 - D_i$ on the transformed outcome $W_i^{1-D} = \mathbb{1}\{Y_i > x\}(1 - D_i)$. Test scores are measured in EYS units.

Figure 7: 2SLS effects by projected grade in year 2

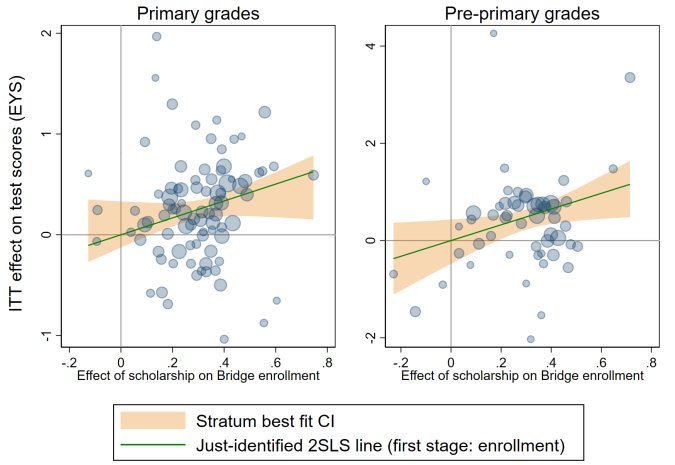


Notes: Figure shows the 2SLS effect of enrolling at Bridge estimated separately by grade. "N" and "P" indicate "nursery" and "preunit" grades. Each sub-figure shows effects for the outcome indicated in the title. The dashed lines indicates 95% confidence intervals. The solid line represents the estimated interaction between the pupil's expected grade in the first year of the study and Bridge enrollment across all grades. All estimates are from specifications that include planned 2016 grade and a linear control for the probability of receiving a scholarship in the randomization strata.

Figure 8: Comparison of ITT effects and first stage effects on fees and Bridge enrollment



(a)



(b)

Notes: Figures show the relationship between the ITT effect on test scores and the first stage effect on fees (top panels) and Bridge enrollment (bottom panels). These are visual-IV plots (Angrist and Pischke, 2009, Chapter 4) for each mediator. To create the plots, the sample is first stratified by county and probable outside option school type. Shaded circles show the ITT and first stage effect for each individual stratum. The size of the circle is inversely proportional to the standard error of the stratum-specific 2SLS estimate. The confidence interval of the best fit line through the circles is given in orange. The red/green lines indicates the pattern that would be expected if the effect of the scholarship were mediated entirely by fees (top panel) or Bridge enrollment (bottom panel). It passes through the origin with slope equal to the estimated 2SLS effect across all strata of fees (top panel in red) and Bridge enrollment (bottom panel in green).

Tables

Table 1: Comparison of Bridge to other Kenyan schools

	Primary grades			Pre-primary grades		
	Public (1)	Bridge (2)	Private (3)	Public (4)	Bridge (5)	Private (6)
Fees and missed class						
Total edu expenditure (KES)	6,061***	13,943	15,611***	4,984***	11,404	11,143
Tuition (KES, incl. meals)	2,230***	8,870	10,213***	2,383***	7,587	7,265
Missed class because unable to pay fees (past term)	0.45***	0.54	0.51	0.47**	0.52	0.56
Average days absent (past term)	2.84	3.19	2.60	1.79***	3.24	2.02**
Pupil-teacher ratio	34^{D}	20^{A}	17^{D}	31^{D}	13^{A}	25^{D}
Length of school day (hours)	8.23***	8.94	8.92	6.08***	7.41	6.78***
School meets on Saturdays	0.28***	0.75	0.43***	0.09***	0.43	0.18***
Teacher characteristics						
More than secondary school education	0.75^{B}	0.23^{A}	0.75^{B}		0.13^{A}	
Certified teacher		0.23^{A}			0.13^{A}	
Average compensation (US$)	563^{C}	116^{B}			116^{B}	
Years experience	17.3^{B}		5.1^{B}			
First year teacher	0.04^{B}		0.10^{B}			
First year teaching at this school	0.19^{B}	0.31^{A}	0.39^{B}		0.19^{A}	
Age	39^{B}	27^{A}	28^{B}		29^{A}	

Notes: This table combines data from multiple sources. External data sources are marked with superscripts. A = Bridge administrative data, B = World Bank Service Delivery Indicators, C = The National Treasury (2018), and D = Ministry of Education (2016a). Items indicated as parent or pupil reports are from data collected by the authors as part of the endline survey described below. For these responses, we restrict the analysis to scholarship non-recipients and include all interviewed subjects (combining the *main*, *brig*, and *priv* samples described below). Total education expenditure, tuition, missed class due to an inability to pay fees are parent reports. The average number of days absent in the past most recent past term, length of school day, and saturday schooling are pupil reports. For pupil and parent reported items, *, **, and *** indicate that the difference between government or private school mean and Bridge mean is statistically significant at the 10%, 5% and 1% level. To compute the average teacher compensation in public primary schools we divide the total disbursement to the Teacher Service Commission (TSC) for primary school level *teacher resource management* by the the total number of primary school teachers employed by the TSC. Teacher compensation in Bridge schools is calculated as the average base salary plus a fifteen percent housing allowance.

Table 2: Description of scholarship application samples and attrition

	P_{main}		PP_{main}		P_{brig}		PP_{brig}		P_{priv}	
	Non-recipient mean (1)	Coef. (2)	Non-recipient mean (4)	Coef. (5)	Non-recipient mean (7)	Coef. (8)	Non-recipient mean (10)	Coef. (11)	Non-recipient mean (13)	Coef. (14)
Panel A: Sample description										
Pupil characteristics										
Female	0.50	0.03*	0.49	0.00	0.48	0.00	0.48	0.02	0.49	0.03
Age	10.56	0.04	5.06	0.10*	10.00	-0.01	5.53	0.32*	9.15	-0.16
Mother is alive	0.90	0.01	0.94	-0.00	0.96	-0.01	0.95	-0.01	0.93	0.03
Father is alive	0.75	-0.01	0.81	-0.01	0.83	0.03	0.84	0.01	0.77	0.01
Home characteristics										
Floor of home is dirt	0.66	-0.02	0.56	0.01	0.55	-0.02	0.45	0.06	0.46	0.04
Walls of home are mud	0.55	-0.02*	0.46	0.02	0.43	-0.01	0.41	0.01	0.33	0.05
Household has electricity	0.37	0.02	0.40	0.00	0.31	0.03	0.40	-0.04	0.46	0.00
Household has latrine	0.76	-0.02	0.72	0.01	0.68	0.02	0.72	0.03	0.73	0.02
HH income (annual 000s KES)	41.90	0.80	47.46	0.09	54.07	-0.62	48.45	-4.65	50.76	-0.44
Primary caregiver characteristics										
Caregiver can read	0.13	-0.01	0.11	-0.01	0.08	0.03**	0.10	-0.00	0.07	0.00
Caregiver can write	0.16	-0.00	0.14	-0.01	0.11	0.02	0.11	-0.01	0.09	-0.01
Caregiver occupation										
Agriculture	0.17	-0.01	0.11	-0.01	0.14	-0.00	0.09	0.01	0.08	-0.02
Casual laborer	0.41	0.00	0.40	0.01	0.36	-0.02	0.42	-0.06	0.38	0.00
Small business owner	0.17	-0.00	0.17	-0.00	0.19	-0.06***	0.12	0.05	0.17	0.02
Observations	2,180	4,621	1,735	3,811	636	1,204	208	388	527	991
F-stat of joint test		1.44		0.99		1.59		1.10		1.21
P-value		0.13		0.47		0.08		0.36		0.26
Panel B: Follow-up rate										
Overall	0.877		0.864		0.866		0.896		0.759	
Recipients	0.870		0.854		0.862		0.874		0.739	
Non-recipients	0.884		0.873		0.871		0.923		0.784	
Differential	0.017		0.013		0.009		0.039		0.049	
P-value	0.06		0.18		0.60		0.14		0.03	
Observations	5,267		4,411		1,390		443		1,305	

Notes: Panel A of this table describes the study sample and compares scholarship recipients and non-recipients in terms of baseline covariates. Sample is restricted to those pupils in the final evaluation sample who were successfully interviewed at endline. The covariates reported in this table are from the application used to enroll the student in the scholarship program. Household income is winsorized at the 99th percentile. The coefficient estimates are from a test of the difference between scholarship recipients and non-recipients. The specification used controls for strata dummies. ***, **, and * indicate significance at 1%, 5%, and 10%. Panel B of the table shows the follow-up rate for each sample. The follow-up differential is estimated using OLS regression including strata dummy controls, and therefore may not equal the difference in mean follow-up rates among scholarship recipients and non-recipients.

Table 3: The effect of winning the scholarship on school enrollment

	P_{main}		PP_{main}		P_{brig}		PP_{brig}		P_{priv}	
	Non-recipient mean (1)	Coef. (2)	Non-recipient mean (3)	Coef. (4)	Non-recipient mean (5)	Coef. (6)	Non-recipient mean (7)	Coef. (8)	Non-recipient mean (9)	Coef. (10)
Panel A: Enrollment in year 2 (2017)										
Bridge	0.166	0.353***	0.278	0.340***	0.616	0.244***	0.673	0.180***	0.307	0.275***
		(0.013)		(0.015)		(0.024)		(0.041)		(0.032)
Public	0.733	-0.316***	0.346	-0.143***	0.281	-0.182***	0.173	-0.094***	0.353	-0.126***
		(0.014)		(0.014)		(0.022)		(0.033)		(0.029)
Private	0.093	-0.036***	0.354	-0.189***	0.102	-0.062***	0.154	-0.086***	0.334	-0.146***
		(0.008)		(0.014)		(0.014)		(0.032)		(0.027)
Unenrolled	0.008	-0.002	0.022	-0.009*	0.000		0.000		0.006	-0.003
		(0.003)		(0.005)						(0.006)
Panel B: Enrollment in year 1 (2016)										
Bridge	0.196	0.374***	0.331	0.364***	0.797	0.135***	0.841	0.070**	0.391	0.227***
		(0.013)		(0.015)		(0.019)		(0.034)		(0.033)
Public	0.712	-0.334***	0.255	-0.116***	0.142	-0.093***	0.067	-0.017	0.264	-0.093***
		(0.014)		(0.012)		(0.017)		(0.025)		(0.027)
Private	0.089	-0.039***	0.346	-0.200***	0.061	-0.042***	0.091	-0.053**	0.338	-0.130***
		(0.008)		(0.014)		(0.011)		(0.026)		(0.027)
Unenrolled	0.003	-0.001	0.068	-0.048***	0.000		0.000		0.008	-0.003
		(0.002)		(0.007)						(0.006)
Panel C: Years of exposure										
Bridge	0.362	0.727***	0.609	0.704***	1.414	0.379***	1.514	0.249***	0.698	0.501***
		(0.025)		(0.028)		(0.038)		(0.069)		(0.062)
Public	1.445	-0.650***	0.601	-0.258***	0.423	-0.275***	0.240	-0.111**	0.617	-0.220***
		(0.026)		(0.024)		(0.035)		(0.053)		(0.052)
Private	0.182	-0.075***	0.700	-0.388***	0.164	-0.105***	0.245	-0.139***	0.672	-0.276***
		(0.014)		(0.026)		(0.022)		(0.053)		(0.050)
Unenrolled	0.011	-0.003	0.090	-0.057***	0.000		0.000		0.013	-0.006
		(0.004)		(0.010)						(0.012)
Number of observations	2,180	4,621	1,735	3,811	636	1,204	208	388	527	991

Notes: Each row represents a separate specification. Sample is restricted to those pupils with endline test scores. The prefix P refers to primary school, and the prefix PP refers to pre-primary. The subscript *brig* indicates the sample of pupils who were continuing Bridge students, and the P_{priv} sample is the sample of pupils who said they would enroll at private primary schools in the absence of the scholarship. All specifications include strata fixed effects. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 4: Effects of scholarship and Bridge enrollment on academic subject knowledge

	Primary			Pre-primary			
	EYS		SD	EYS		SD	EYS
	ITT	2SLS	2SLS	ITT	2SLS	2SLS	gradient
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Math & language scores							
Subject knowledge index	0.31***	0.89***	0.81***	0.51***	1.48***	1.35***	0.92***
	(0.04)	(0.12)	(0.11)	(0.06)	(0.16)	(0.15)	(0.01)
Higher order skills index	0.29***	0.80***	0.90***	0.28***	0.82***	0.91***	1.11***
	(0.04)	(0.11)	(0.12)	(0.05)	(0.15)	(0.16)	(0.01)
Language knowledge index	0.33***	0.94***	0.61***	0.66***	1.91***	1.23***	0.65***
	(0.05)	(0.14)	(0.09)	(0.07)	(0.21)	(0.13)	(0.01)
Math score	0.30***	0.85***	0.90***	0.41***	1.20***	1.27***	1.06***
	(0.04)	(0.12)	(0.13)	(0.05)	(0.15)	(0.16)	(0.01)
English score	0.41***	1.16***	0.68***	0.65***	1.89***	1.11***	0.59***
	(0.05)	(0.14)	(0.08)	(0.07)	(0.20)	(0.12)	(0.01)
Kiswahili score	0.26***	0.74***	0.45***	0.66***	1.93***	1.18***	0.61***
	(0.06)	(0.16)	(0.10)	(0.08)	(0.22)	(0.14)	(0.01)
Observations		4,587			3,805		
Panel B: Science & social studies							
Science & social studies	0.48***	1.35***	0.72***	0.09	0.29	0.15	0.53***
	(0.07)	(0.19)	(0.10)	(0.08)	(0.23)	(0.12)	(0.01)
Science	0.33***	0.92***	0.77***	0.05	0.16	0.13	0.83***
	(0.06)	(0.17)	(0.14)	(0.08)	(0.23)	(0.19)	(0.01)
Social studies	0.52***	1.47***	0.58***	0.09	0.29	0.12	0.39***
	(0.07)	(0.19)	(0.08)	(0.08)	(0.23)	(0.09)	(0.01)
Observations		4,526			794		
Local content score	0.54***	1.45***	0.59***				0.41***
	(0.08)	(0.21)	(0.09)				(0.01)
Observations		3,750					

Notes: The standard deviation unit meaasures test score outcomes in standard deviations of the Standard 4 test score distribution. All specifications include controls for baseline characteristics, planned 2016 grade, and strata dummies. The higher order skills index restricts excludes subject test items that reflect remembering/recalling information. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 5: Distributional effects of Bridge enrollment

	Primary (1)	Pre-primary (2)
Panel A: Quantile treatment effects		
Q(0.10)	0.97***	1.72***
	(0.21)	(0.28)
Q(0.20)	0.90***	1.78***
	(0.17)	(0.23)
Q(0.80)	0.82***	1.17***
	(0.16)	(0.20)
Q(0.90)	0.74***	1.15***
	(0.19)	(0.26)
Q(0.20)-Q(0.80)	0.08	0.61***
	[0.08]	[0.11]
Q(0.10)-Q(0.90)	0.24***	0.58***
	[0.10]	[0.15]
Panel B: Test of first order stochastic dominance		
p-value	0.004	0.008
Panel C: Scholarship effect on test score dispersion		
SD recipients	1.42	1.72
SD non-recipients	1.49	1.79
p-value (F-test)	0.014	0.091
Observations	4,621	3,811

Notes: Primary = P_{main} and Pre-primary = P_{main}. All tests are reported in EYS units. Quantile effects are from the method of Abadie et al. (2002). The specification for quantile regression includes planned 2016 grade and linear control for probability of treatment assignment. Quantile analytical standard errors are reported in parentheses. Bootstrap standard errors of the difference in quantile effects are reported in brackets and represent that standard deviation of the statistic from 1000 bootstrap samples. The p-value for the test of first order stochastically dominance is the result of a test from Abadie (2002) where low values support first order stochastic dominance. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 6: Variation in Bridge enrollment effects across locations

(a) Sargan test for heterogeneous academy effects

	Primary grades (1)	Pre-primary grades (2)	Pooled (3)
Subject knowledge index			
Sargan	317.9	323.9	321.0
DF	321	324	357
p-value	0.538	0.492	0.914
N pupils	4,621	3,811	8,432
N Academies	348	350	368

Notes: This table reports results from a Sargan over-identification test using interaction of scholarship assignment with location (academy) indicators. Primary = P_{main} and Pre-primary = PP_{main}. All specifications include controls for the baseline projected grade in 2016, academy dummies and the probability of treatment for the strata.

(b) Random coefficients estimates

		Primary		Pre-primary	
		Estimate (1)	Standard error (2)	Estimate (3)	Standard error (4)
$E[\alpha_{1j}]$	Mean Bridge outcome	-1.387	0.214	-2.214	0.255
$E[\alpha_{0j}]$	Mean non-Bridge outcome	-2.297	0.137	-3.653	0.233
$E[\alpha_{1j} - \alpha_{0j}]$	Mean Bridge enrollment effect	0.909	0.151	1.439	0.173
$\sqrt{(Var[\alpha_{1j} - \alpha_{0j})}$	Std. dev. of Bridge enrollment effects	0.339	0.048	0.282	0.056
$\sqrt{(Var(\alpha_{1j}))}$	Std. dev. of Bridge site means	0.253	0.028	0.422	0.054
$\sqrt{(Var(\alpha_{0j}))}$	Std. dev. of non-Bridge site means	0.352	0.030	0.328	0.041
$Cov(\alpha_{1j}, \alpha_{0j})$	Cov. of Bridge & non-Bridge site means	0.036	0.021	0.103	0.034

Notes: Table reports estimates from the random coefficients model from (Walters, 2015). Sample restricts to those pupils in the *main* sample. The maximum simulated likelihood procedure uses 1,000 simulations for each Bridge location. All specifications control linearly for the randomization strata probability of scholarship assignment and indicators for projected grade. In the pre-primary sample, Baby class is the omitted grade category. In the primary school sample, Standard 4 is the omitted category.

Table 7: Effects of scholarship and Bridge enrollment on timely grade progression

		Primary		Pre-primary	
		ITT	2SLS	ITT	2SLS
		(1)	(2)	(3)	(4)
Timely grade progression					
... in 2017	Coef.	0.07***	0.20***	0.06***	0.19***
	SE	(0.01)	(0.04)	(0.01)	(0.04)
	N	4,589		3,777	
	Mean	0.74		0.73	
... in 2019	Coef.	0.05***	0.14***	0.06***	0.15***
	SE	(0.02)	(0.04)	(0.02)	(0.05)
	N	3,767		3,022	
	Mean	0.64		0.68	

Notes: Table shows the effects of the scholarship and Bridge enrollment on measures of academic achievement. Primary = P_{main} and Pre-primary = PP_{main}. *Timely grade progression* indicates that the pupil reached at least the grade that they would have achieved given the grade in which they planned to enroll in year 1 (2016). All specifications include controls for baseline characteristics, planned 2016 grade, and randomization strata fixed effect. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 8: Heterogeneous 2SLS effects of attending Bridge at cost on subject knowledge index

	Primary			Pre-primary		
Covariate	Attended Bridge (1)	× Cov (2)	Cov (3)	Attended Bridge (4)	× Cov (5)	Cov (6)
Panel A: Pupil characteristics						
Expected achievement index	0.823***	-0.356**	0.313***	1.480***	-0.010	0.174
	0.119	0.167	0.092	0.161	0.277	0.166
		4,621			3,811	
Pupil is female	0.884***	-0.030	0.193**	1.569***	-0.209	0.372**
	0.167	0.211	0.089	0.204	0.317	0.162
		4,621			3,811	
Panel B: Household characteristics						
Floor of home is dirt	0.641***	0.321	-0.228**	1.230***	0.402	-0.326**
	0.207	0.236	0.098	0.268	0.326	0.165
		4,621			3,811	
Household has electricity	0.889***	-0.075	0.250***	1.625***	-0.388	0.347**
	0.147	0.212	0.101	0.205	0.318	0.171
		4,621			3,811	
HH income (annual 000s KES)	0.872***	0.001	0.000	1.470***	-0.004	0.002
	0.117	0.002	0.001	0.160	0.003	0.001
		4,621			3,811	
Panel C: Location characteristics						
Mean KCPE w/in 2km	0.865***	0.034	-0.015	1.481***	0.038	0.064
	0.117	0.104	0.049	0.161	0.166	0.088
		4,621			3,811	
Rural location	1.162***	-0.456**	0.015	1.323***	0.317	-0.383***
	0.200	0.231	0.099	0.224	0.316	0.162
		4,621			3,811	
Class size (projected if enrolls in Bridge)	1.048***	-0.008	0.001	1.391***	0.003	-0.009
	0.267	0.016	0.007	0.358	0.018	0.010
		3,729			3,206	
Panel D: Bridge teacher characteristics						
More than 2 years experience	0.730***	0.113	-0.057			
	(0.188)	(0.196)	(0.078)			
		13,101				
Years experience teaching at Bridge	0.592***	0.090	-0.036			
	(0.223)	(0.075)	(0.030)			
		13,101				
KCSE score (standardized)	0.750***	-0.073	0.040			
	(0.132)	(0.105)	(0.042)			
		8,693				

Notes: Table reports the heterogeneous effects of enrolling at Bridge across pupil, household, location, and prospective Bridge teacher characteristics. Primary = P_{main} and Pre-primary = PP_{main}. Test score outcomes are measured in EYS units. All specifications include baseline characteristics, planned 2016 grade, and control linearly for the randomization strata probability of scholarship assignment. The unit of observation is a pupil in Panels A through D. Panel D stacks multiple individual subject scores to leverage within pupil variation in characteristics of teachers for specific subjects. Data on teacher characteristics are only available for primary school grades, so we only examine effects in the primary school sample. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 9: Effects of scholarship and Bridge enrollment on primary school completion outcomes

		Primary				
		Pooled		Standard 5	Standard 6	Standard 7
		ITT	2SLS	2SLS	2SLS	2SLS
		(1)	(2)	(3)	(4)	(5)
Took KCPE	Coef.	0.06***	0.15***	0.25**	0.33***	-0.08
	SE	(0.02)	(0.06)	(0.11)	(0.10)	(0.09)
	N	1,616		503	490	623
	Mean	0.74		0.57	0.76	0.86
KCPE score	Coef.	5.9**	16.2**	13.5	-3.6	40.0***
	SE	(3.0)	(8.2)	(16.8)	(15.1)	(12.8)
	N	1,293		338	424	531
	Mean	260.8		265.2	266.4	254.0
Passed KCPE	Coef.	0.06***	0.17***	0.18	0.19	0.18
	SE	(0.03)	(0.07)	(0.11)	(0.13)	(0.12)
	N	1,616		503	490	623
	Mean	0.41		0.32	0.45	0.46
Passed KCPE (conditional on taking)	Coef.	0.04	0.12	0.14	-0.03	0.26**
	SE	(0.03)	(0.08)	(0.15)	(0.14)	(0.12)
	N	1,293		338	424	531
	Mean	0.56		0.57	0.58	0.53
Transitioned to secondary school	Coef.	0.01	0.02		-0.04	0.05
	SE	(0.03)	(0.07)		(0.13)	(0.08)
	N	1,077			523	554
	Mean	0.71			0.58	0.83

Notes: Table shows the effects of the scholarship and Bridge enrollment on primary school completion outcomes. Primary = P_{main} and Pre-primary = PP_{main}. The sample for KCPE outcomes restricts to pupils who were planning to enter Standards 5 or higher in 2016. The standard deviation of KCPE scores for the sub-sample of non-recipients is 58. All tests include baseline characteristics, planned 2016 grade, and control linearly for the randomization strata probability of scholarship assignment, since strata controls lead to sample loss due to a lack of variation in some strata cells. *Mean* is the non-recipient mean. We treat KCPE test taking status as missing if no response was received on any of the surveys after the year in which the pupil was projected to take the KCPE. *Transitioned to secondary school* indicates that the pupil was observed to be enrolled in secondary school either during the endline pupil survey at the end of 2017 or in the 2019 phone call tracking survey. The sample for *secondary school transition* restricts to pupils who were planning to enter Standards 6 or higher in 2016. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 10: Effects of scholarship and Bridge enrollment on cognitive and non-cognitive outcomes

	Primary			Pre-primary			
	EYS		SD	EYS		SD	EYS
	ITT	2SLS	2SLS	ITT	2SLS	2SLS	gradient
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Raven's matrices	0.09	0.27	0.06	0.15	0.44	0.09	0.21***
	(0.15)	(0.41)	(0.09)	(0.11)	(0.31)	(0.07)	(0.00)
		4,591			3,794		
Digit span recall	0.44***	1.23***	0.21***	0.61***	1.76***	0.30***	0.17***
	(0.17)	(0.46)	(0.08)	(0.17)	(0.49)	(0.08)	(0.01)
		4,541			3,452		
Draw a picture task				0.38***	1.07***	0.37***	0.34***
				(0.16)	(0.44)	(0.15)	(0.05)
					2,919		
Head-knees task				0.22*	0.64*	0.20*	0.32***
				(0.13)	(0.38)	(0.12)	(0.05)
					2,872		
Receptive vocabulary							
All items	0.10	0.31	0.17	0.42***	1.21***	0.65***	0.54***
	(0.11)	(0.35)	(0.19)	(0.07)	(0.20)	(0.11)	(0.02)
		1,246			3,743		
English items	0.12	0.40	0.17	0.50***	1.45***	0.62***	0.43***
	(0.14)	(0.45)	(0.19)	(0.08)	(0.23)	(0.10)	(0.01)
		1,246			3,743		
Kiswahili items	0.07	0.22	0.11	0.33***	0.95***	0.49***	0.51***
	(0.12)	(0.37)	(0.19)	(0.08)	(0.23)	(0.12)	(0.02)
		1,246			3,743		
Spoons task							
Unusual uses	0.63	1.84	0.07	-0.33	-1.00	-0.04	0.04***
	(0.86)	(2.48)	(0.09)	(0.81)	(2.46)	(0.09)	(0.01)
		3,615			2,887		
Total uses	-0.02	-0.05	-0.00	-0.14	-0.44	-0.04	0.08***
	(0.37)	(1.07)	(0.09)	(0.38)	(1.16)	(0.10)	(0.01)
		3,615			2,887		

Notes: For the standard deviation unit, effects on *Raven's matrix*, *digit span recall* and *spoons task* outcomes expressed in standard deviations of the Standard 4 test score distribution. Effects on *draw a picture task*, *head-knees task* are expressed SD units of the preunit distribution. *Receptive vocabulary* outcomes are expressed in SD units of the Standard 2 distribution. All specifications include controls for baseline characteristics, planned 2016 grade, and strata dummies. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 11: Effects of Bridge enrollment on intermediate inputs/mediators and classroom experience

	Primary				Pre-primary			
	Non-recipient mean	ITT	2SLS	N	Non-recipient mean	ITT	2SLS	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Length of day (hours)	8.62	0.21***	0.59***	3,746	6.71	0.61***	1.79***	2,356
		(0.06)	(0.16)			(0.10)	(0.27)	
Attend Saturday school every week	0.24	0.14***	0.40***	4,514	0.10	0.08***	0.24***	3,623
		(0.01)	(0.03)			(0.01)	(0.03)	
Teacher never absent from class	0.36	0.05***	0.14***	4,427	0.49	-0.01	-0.03	3,107
		(0.01)	(0.04)			(0.02)	(0.05)	
Teacher uses local language	0.30	-0.07***	-0.19***	4,547	0.29	-0.04***	-0.11***	3,593
		(0.01)	(0.04)			(0.01)	(0.04)	
Number textbooks	3.53	0.23***	0.65***	4,534	2.03	0.17***	0.48***	3,425
		(0.06)	(0.15)			(0.07)	(0.19)	
Days with homework (usual week)	3.37	0.59***	1.67***	4,550	3.60	0.34***	0.97***	3,725
		(0.05)	(0.14)			(0.06)	(0.17)	
Study hours (usual day)	1.45	0.08*	0.20*	2,957				
		(0.05)	(0.12)					
Taken test similar to endline test	0.81	0.03***	0.09***	4,475	0.77	0.02	0.06	806
		(0.01)	(0.03)			(0.04)	(0.10)	
Instructional adaptation index	0.00	0.14***	0.40***	4,162	0.03	0.07***	0.20***	3,034
		(0.02)	(0.06)			(0.03)	(0.07)	
Classroom crowding index	-0.12	-0.15***	-0.41***	4,541	0.21	-0.06*	-0.17*	3,691
		(0.03)	(0.08)			(0.03)	(0.10)	
Classroom engagement index	-0.01	0.06**	0.16**	4,488	-0.07	0.01	0.04	879
		(0.03)	(0.08)			(0.10)	(0.29)	
Parental engagement index	-0.00	0.14***	0.42***	4,559	0.05	0.23***	0.67***	3,666
		(0.03)	(0.09)			(0.04)	(0.12)	
Test preparation index	0.01	0.09***	0.24***	4,478	-0.10	0.07	0.23	807
		(0.03)	(0.09)			(0.08)	(0.24)	
Field has hazards	0.34	0.03*	0.08*	4,169	0.39	0.01	0.04	3,321
		(0.01)	(0.04)			(0.02)	(0.05)	
School has wall or fence	0.66	0.02	0.06	4,532	0.64	0.05***	0.15***	3,642
		(0.01)	(0.04)			(0.01)	(0.04)	
Teachers carry sticks or canes	0.81	-0.05***	-0.14***	4,548	0.79	-0.01	-0.03	3,689
		(0.01)	(0.04)			(0.01)	(0.04)	
Teachers hit, slap or pinch	0.83	-0.02*	-0.06*	4,549	0.80	0.00	0.00	3,687
		(0.01)	(0.03)			(0.01)	(0.04)	

Notes: Table reports the effects of the scholarship and enrolling at Bridge on pupil and caregiver reported perceptions of inputs or mediators. Primary = P_{main} and Pre-primary = PP_{main}. Construction of indices is described in the text. School day length excludes Saturday schooling. Length of school day is calculated based on pupil responses to school start and end times, where implausible cases outside one to eleven are set to missing. The sample size for the non-recipient mean is approximately 0.5 times of N for all outcomes. All specifications include baseline characteristics, planned 2016 grade, and strata fixed effect. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 12: Effects of scholarship on education expenditure and missed class

	Primary			Pre-primary		
	Non-recipient Mean (1)	ITT (2)	N (3)	Non-recipient Mean (4)	ITT (5)	N (6)
Panel A: Parent reported outcomes						
Total edu expenditure (000s KES)	7.50	-1.27***	4,498	8.74	-2.49***	3,670
		(0.22)			(0.24)	
School fees (000s KES)	3.69	-1.35***	4,327	5.42	-2.32***	3,575
		(0.15)			(0.16)	
Missed class b/c unable to pay fees	0.47	-0.19***	4,268	0.52	-0.26***	3,523
		(0.01)			(0.02)	
Tutoring expenditure (000s KES)	0.53	-0.03	4,378	0.36	-0.07*	3,606
		(0.04)			(0.04)	
Any tutoring	0.31	-0.00	4,439	0.18	0.01	3,581
		(0.01)			(0.01)	
Number tutoring hrs per usual wk	2.46	-0.05	4,439	1.20	0.14	3,581
		(0.14)			(0.12)	
Panel B: Pupil reported outcomes						
Missed class b/c unable to pay fees	0.12	-0.05***	4,563	0.10	-0.03***	3,530
		(0.01)			(0.01)	
Number days absent (past term)	2.13	-0.37***	4,548	1.94	-0.53***	3,450
		(0.13)			(0.16)	
>5 days absent past term	0.08	-0.02**	4,548	0.06	-0.01*	3,450
		(0.01)			(0.01)	
>10 days absent past term	0.03	-0.01*	4,548	0.03	-0.01***	3,450
		(0.01)			(0.01)	

Notes: Primary = P_{main} and Pre-primary = P_{main}. Sample restricts to those pupils with non-missing test scores. All specifications include baseline controls, planned 2016 grade and strata controls. Values for pupil absences that exceed 30 days are set to 30 days, approximately the 99th percentile. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 13: 2SLS estimates with multiple endogenous variables

	Primary grades			Pre-primary grades		
	OLS	2SLS		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Separate estimation of Bridge enrollment and fee effects						
Enrolled at Bridge	0.52***	0.95***	1.06***	0.47***	1.90***	2.01***
	(0.06)	(0.16)	(0.15)	(0.09)	(0.43)	(0.36)
		[114.28]	[38.56]		[26.37]	[11.95]
School fees paid (000s KES)	0.04***	0.03*	0.04***	0.03***	0.07	0.08**
	(0.00)	(0.02)	(0.02)	(0.01)	(0.05)	(0.04)
		[43.58]	[14.48]		[14.67]	[6.14]
Kleibergen-Paap F-stat		25.89	10.41		7.32	3.89
Observations		6,112	6,441		3,726	3,950
Panel B: Separate estimation of public and non-Bridge private effects						
Bridge relative to public school	0.27***	0.78***	0.96***	0.34***	1.07	2.10***
	(0.08)	(0.14)	(0.13)	(0.11)	(1.45)	(0.68)
		[156.28]	[57.62]		[2.80]	[3.92]
Bridge relative to other private schools	0.05	0.89*	0.01	0.23**	1.80*	1.10**
	(0.10)	(0.52)	(0.49)	(0.10)	(1.08)	(0.56)
		[16.67]	[5.80]		[5.28]	[6.44]
Kleibergen-Paap F-stat		11.39	4.82		1.05	1.71
Observations		5,305	5,612		3,575	3,811
Instruments						
Academy-intention prediction interactions		X			X	
Intention and demographic interactions			X			X

Notes: This table reports estimates of specifications with multiple endogenous variables. All estimates are reported in EYS units. Panel A reports results from specifications that separately estimate the effects of enrolling at Bridge and of paying fees. School fees in Panel A include tuition, meals, and uniform costs. Panel B reports results estimates of the effect of enrolling at Bridge relative to public schools and other private schools. Column (1) reports a non-experimental OLS specification restricting the sample to the control group. All 2SLS specifications are identified using multiple instruments formed by interacting scholarship assignment with a vector of baseline controls. *Academy-intention predictions* are four instruments representing the empirical likelihood that a pupil in a particular academy-intention cell will enroll in a public school, Bridge school, non-Bridge private school, or remain unenrolled in the absence of the scholarship. Each prediction is the empirical Bayes shrunk leave-i-out share of pupils choosing each enrollment type among scholarship non-recipients. The *intention and demographic interactions* include the following ten indicators: being enrolled in any school at time of application; being enrolled in NGO/private school; any prior Bridge enrollment; having plans to enroll in a Bridge school/public school/non-Bridge private school in the absence of the scholarship; primary caregiver's is a agricultural/casual laborer; whether the household applied through IPA; and Inverse Hyperbolic Sine (IHS) transformation of reported household monthly income. All specifications include baseline characteristics, planned 2016 grade, and control linearly for the randomization strata probability of scholarship assignment. 2SLS specifications use instruments specified at the bottom of the table. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Appendices

A2

SHINE Fluency_L1

Materials Needed: LM Literacy textbooks, Exercise books

Note: Letters A-I (AY) make the A sound as in 'ate'

Sound Combinations – 5 minutes

1. **Copy as I write.**
2. Write on board:

aim
rain
stain
paint
sprain

3. **Eyes on me.** Scan.
4. **The letters A—I go together and usually make the sound AY, as in AIM.**
5. **Say AY.** [Signal] *AY*
6. **You will read the words that have letters A—I.**
7. **Say the sound for the underlined part, then read the word.**
8. **Touch word 1. What sound?** [Signal] *AY*
9. **What word?** [Signal] *Aim*
10. **Next word. What sound?** [Signal] *Ay*
11. **What word?** [Signal] *Rain*

12. Repeat last 2 lines for each word.
13. **Copy as I write.**
14. Add to board:

aimless
grain
faint
plains
raining

15. **Touch word 1. What word?** [Signal] *Aimless*
16. **Touch word 2. What word?** [Signal] *Grain*
17. Repeat last line for each word.

Build Ups – 5 minutes

18. Clean board and write:

lain

19. **Eyes on me.** Scan.
20. **What word?** [Signal] *Lain*
21. Change word:

plain

22. **What word now?** [Signal] *Plain*
23. Change word:

plains

24. **What word now?** [Signal] *Plains*

50. **Say the whole word.** [Signal] *Confident*
51. **Touch word 2.**
52. **Word 2 is instrument. What word?** [Signal] *Instrument*
53. **I will say 'instrument' and pause between the parts.**
54. **The parts are in** (pause) **stru** (pause) **ment.**
55. **What's the first part?** [Signal] *Ins*
56. **The middle part is spelled T-R-U.**
57. **What does that part say in the word?** [Signal] *tru*
58. **What's the last part?** [Signal] *ment*
59. **Say the whole word.** [Signal] *Instrument*
60. **Touch word 3.**
61. **Word 3 is audience. What word?** [Signal] *Audience*
62. **I will say 'audience' and pause between the parts.**
63. **The parts are au** (pause) **di** (pause) **ence.**
64. **What's the first part?** [Signal] *au*
65. **The middle part is spelled D-I.**
66. **What does that part say in the word?** [Signal] *di*
67. **What's the last part?** [Signal] *ence*
68. **Say the whole word.** [Signal] *Audience*
69. **Touch word 4.**
70. **Word 4 is awarded. What word?** [Signal] *Awarded*

A4

71. **I will say 'awarded' and pause between the parts.**
72. **The parts are a** (pause) **ward** (pause) **ed.**
73. **What's the first part?** [Signal] *a*
74. **The middle part is spelled W-A-R-D.**
75. **What does that part say in the word?** [Signal] *ward*
76. **What's the last part?** [Signal] *ed*
77. **Say the whole word.** [Signal] *Awarded*
78. **Eyes on me.**
79. Touch word 1. **What word?** [Signal] *Confident*
80. Touch word 2. **What word?** [Signal] *Instrument*
81. Touch word 3. **What word?** [Signal] *Audience*
82. Touch word 4. **What word?** [Signal] *Awarded*
83. **Copy as I write.**
84. Add to board:

1. The mayor awarded him a prize.
2. The audience clapped after the concert.
3. A piano is a kind of instrument.
4. She is confident she will pass her maths exam.

85. When pupils finish copying, **Eyes on me.**
86. **These sentences use the words you just read.**
87. **Read sentence 1.** [Signal] *The mayor awarded him a prize.*

88. **Read sentence 2.** [Signal] *The audience clapped after the concert.*
89. **Read sentence 3.** [Signal] *A piano is a kind of instrument.*
90. **Read sentence 4.** [Signal] *She is confident she will pass her maths exam.*

Vocabulary – 5 minutes

91. **We will learn what these words mean.**
92. **Touch word 1. What word?** [Signal] *Confident*
93. **If you feel confident, you feel sure.**
94. **What is another way to say: I feel sure I will earn top marks?** [Signal] *I feel confident I will earn top marks.*
95. **I feel confident I will earn top marks. Again.** [Signal] *I feel confident I will earn top marks.*
96.
97. **Touch word 2. What word?** [Signal] *Awarded*
98. **If you gave someone a prize, you awarded them.**
99. **What is another way to say: He gave me a prize?** [Signal] *He awarded me a prize.*
100. **He awarded me a prize. Again.** [Signal] *He awarded me a prize.*
101.

102. **Touch word 3. What word?** [Signal] *Instrument*
103. **We use instruments to play music. A drum is a kind of instrument.**
104. **What is another way to say: She plays a musical tool?** [Signal] *She plays an instrument.*
105. **She plays an instrument. Again.** [Signal] *She plays an instrument.*
106.
107. **Touch word 4. What word?** [Signal] *Audience*
108. **The people who watch a performance, like a play or concert, are the audience.**
109. **What is another way to say: The people watching liked the film?** [Signal] *The audience liked the film.*
110. **The audience liked the film. Again.** [Signal] *The audience liked the film.*

Fluency Read – 12 minutes

111. Pass out / tell pupils to open Literacy LM textbooks.
112. **Turn to page 5.**
113. **Point to the first full stop.** Scan.
114. **What word comes before the full stop?** [Signal] *Town*
115. **When we read aloud, we pause at a full stop.**

116. **I will read paragraph 1. Notice how I pause at the full stops.**

117. **Follow as I read.**

118. **Emmanuel was a boy from a small town.** (pause)

119. **He was shy and did not have many friends.** (pause)

120. **His favourite instrument was the piano.** (pause)

121. **He spent hours alone learning new music.** (pause)

122. **Your turn to read and pause at the full stops.**

123. **Read paragraph 2. Begin.** [Signal]

124. *Everyone in the town thought Emmanuel was rude because he did not talk a lot.* (pause)

125. *But Emmanuel was only shy.* (pause)

126. *He was confident when he played music, but not when he had to talk.* (pause)

127. *Emmanuel wanted to join a choir or a band with other people so he could make friends.* (pause)

128. **Stop.**

129. **You will finish the text with your deskmates. Each pupil reads 1 paragraph. Remember to pause at the full stops.**

130. **If you finish early, start again from the beginning.**

131. **You have 7 minutes. Begin.**

132. Check, respond, leave:

Ensure pupils:

- pause at full stops
- pronounce the following words correctly: awarded, audience, instruments

133. After 7 minutes, **Eyes on me.**

134. **Our fluency class is finished.**

135. Lead a cheer.

Appendix B Appendix tables

Table A1: Data sets

Name	Date	Description	Key information collected
Application data	Nov 2015	The application data comprised information on the pupils' demographics, prior education, caregiver demographics and contact information	current and planned school type, household assets, caregiver occupation, previous assessment scores
First phone call survey	Jun 2016	The objective was to maintain up-to-date contact information to facilitate the endline survey	school type attended at in first year
Second phone call survey	Jun 2017	The objective was to maintain up-to-date contact information to facilitate the endline survey	school type attended at in second year
Endline survey	Nov 2017	This comprised home visits and collected data on the main outcomes from both pupils and caregivers	subject knowledge, cognitive and non-cognitive achievement, classroom experience, parental engagement
Third phone call survey	Jun 2019	The objective was to follow-up with pupils in 2019	KCPE performance, grade progression, secondary school transition

Table A2: Indices constructed from pupil and caregiver reports

Index	Description	Pre-specified?
Local content knowledge	Based on a subset of science and social studies items that represent local content specific to Kenya/East Africa.The items were identified by the research team and the final score was computed using IRT (see PAPSA). Only for standard 3 and above.	Yes
Instructional adaptation	The pupil survey asked "Last year, if you were falling behind in class, did your teachers or school give you extra help for example extra classes, or different classes or readings or exercises?" and "Last year, if you were ahead of your classmates, did your teachers or school give you extra help for example extra classes, or different classes or readings or exercises?" Caregivers were asked the same questions about the pupil's school. The two responses on both the pupil and caregiver surveys are combined into a single index by standardizing the items and taking the average of the standardized responses. The result is standardized to the distribution of scholarship non-recipients in the *main* samples.	Yes
Higher-order skills	Same as subject knowledge index but excludes items that may reflect recalling information or rote memorization. Items identified prior to analysis. Final score was computed using IRT.	No
Classroom crowding	This index combines pupil responses on four items: 1) "Were there any children who did not have a desk or chair in your classes?", 2) "In your classroom at [school], did some pupils bother others?", 3) "Did some pupils fight?", and 4) "Did outside noises make it hard to hear your teacher?". Items are signed so that more positive values indicate more crowded classrooms. Same standardization procedure as *instructional adaptation index*	No
Teacher-classroom engagement	Combines six items from the pupil survey: 1) "if you didn't understand something, your teachers explained it another way", 2) "your teacher asked questions to be sure you were following along", 3) "how often did teachers ask questions to students", 4) "how often did students ask questions to teachers", 5) "teachers only asked certain students questions", and 6) "when my teachers were grading my work, they wrote on my papers to help me understand". Items were only asked for pupils in primary school, so only the oldest cohort of the pre-primary sample have this outcome. Same standardization procedure as *instructional adaptation index*	No
Test preparation	Combines two items from the pupil survey that are intended to capture the amount of practice that pupils have received prior to the test: 1) "You spent a lot of time practicing taking exams" and 2) "a lot of class time was spent on getting you prepared to take a test". Same standardization procedure as *instructional adaptation index.*	No
Parental engagement	This index combines four responses from the caregiver survey: 1) an indicator for whether someone in the household is on the pupil's school committee, 2) an indicator for whether the caregiver knows the pupil's class rank, 3) the number of times the caregiver reported meeting with the pupil's teacher in the past year, and 4) the number of times the caregiver reported meeting the head teacher in the past year. Same standardization procedure as *instructional adaptation index.*	No

Table A3: Effects of scholarship and Bridge attendance on specific items - Pre-primary only

	PP_{main}			By projected 2017 grade								
				Nursery			Pre-unit			Standard 1		
	Non-recipient mean	ITT	2SLS	Non-recipient mean	ITT	2SLS	Non-recipient mean	ITT	2SLS	Non-recipient mean	ITT	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Letters per minute	15.80	4.15***	11.90***	13.45	4.60***	12.66***	19.94	3.40***	11.09***	21.73	2.58	7.28
		(0.55)	(1.56)		(0.66)	(1.78)		(1.21)	(3.89)		(1.89)	(5.22)
		1,353	2,938		895	1,974		340	718		118	246
Pupil can...												
...read the word Thursday	0.12	0.05***	0.14***	0.05	0.00	0.00	0.11	0.05*	0.17**	0.31	0.16***	0.44***
		(0.01)	(0.03)		(0.01)	(0.03)		(0.03)	(0.08)		(0.03)	(0.10)
		1,723	3,786		940	2,065		402	845		381	876
...read a simple English sentence	0.22	0.10***	0.30***	0.09	0.06***	0.16***	0.24	0.13***	0.44***	0.51	0.20***	0.55***
		(0.01)	(0.04)		(0.01)	(0.04)		(0.03)	(0.12)		(0.03)	(0.10)
		1,723	3,786		940	2,065		402	845		381	876
...read a simple Kiswahili sentence	0.11	0.06***	0.16***	0.04	0.01	0.02	0.09	0.09***	0.29***	0.32	0.15***	0.41***
		(0.01)	(0.03)		(0.01)	(0.03)		(0.02)	(0.08)		(0.04)	(0.10)
		1,673	3,683		932	2,048		396	832		345	803
...count eight stars	0.83	0.04***	0.13***	0.79	0.06***	0.16***	0.91	0.01	0.03			
		(0.01)	(0.04)		(0.02)	(0.05)		(0.02)	(0.07)			
		1,235	2,692		895	1,974		340	718			
...subtract 5 from 8 stars	0.42	0.05***	0.14***	0.37	0.04*	0.11*	0.58	0.07*	0.21*			
		(0.02)	(0.06)		(0.02)	(0.06)		(0.04)	(0.12)			
		1,235	2,692		895	1,974		340	718			
...add 4+2	0.50	0.07***	0.19***	0.34	0.06***	0.17***	0.62	0.08***	0.27***	0.80	0.06***	0.17***
		(0.01)	(0.04)		(0.02)	(0.06)		(0.03)	(0.11)		(0.03)	(0.07)
		1,673	3,683		932	2,048		396	832		345	803
...complete sequence 9,8,7,__	0.19	0.04***	0.12***	0.10	0.01	0.03	0.15	0.06**	0.19**	0.47	0.10***	0.28***
		(0.01)	(0.04)		(0.01)	(0.04)		(0.03)	(0.09)		(0.04)	(0.10)
		1,673	3,683		932	2,048		396	832		345	803

Notes: Table reports estimates of the effect of the scholarship and Bridge attendance on performance on specific items. *Letters per minute* indicates the number of correct letters identified in one minute. Pupils were shown 60 letters that included both upper and lower case letters. Some letters were repeated. Only 1.3 percent of pupils were able to identify 60 letters in 60 seconds. The simple English sentence was *Ken has a big dog.* The simple Kiswahili sentence was *Kaka alikunywa sharubati.* All specifications include controls for the probability of treatment for the pupil's randomization strata. Note that the predicted outcome for the ability to read a simple sentence in Standard 1 (Column 12) is inside the unit interval. Using the approach of Abadie (2002), we estimate the expected $Y(1)$ for compliers to be 0.82, and $Y(0)$ to be 0.27. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A4: Lee bounds for effect on aggregate subject knowledge index

	ITT		2SLS	
	Lower	Upper	Lower	Upper
	(1)	(2)	(3)	(4)
Panel A: Main sample				
P_{main}				
	0.25***	0.40***	0.71***	1.10***
	(0.04)	(0.04)	(0.11)	(0.11)
	[0.18 , 0.47]		[0.53 , 1.28]	
	N=4,576		N=4,542	
PP_{main}				
	0.42***	0.58***	1.21***	1.66***
	(0.05)	(0.05)	(0.15)	(0.16)
	[0.34 , 0.66]		[0.96 , 1.92]	
	N=3,775		N=3,769	
Panel B: Additional samples				
PP_{brig}				
	-0.20*	0.06	-0.99	0.29
	(0.12)	(0.12)	(0.65)	(0.56)
	[-0.40 , 0.26]		[-2.06 , 1.21]	
	N=379		N=373	
P_{brig}				
	0.07	0.14***	0.30	0.55***
	(0.06)	(0.06)	(0.23)	(0.23)
	[-0.03 , 0.24]		[-0.08 , 0.93]	
	N=1,198		N=1,195	
P_{priv}				
	-0.08	0.35***	-0.29	1.21***
	(0.09)	(0.09)	(0.36)	(0.32)
	[-0.23 , 0.50]		[-0.88 , 1.74]	
	N=965		N=930	

Notes: Subject knowledge index is in EYS units. All specifications include baseline characteristics, planned 2016 grade, and randomization strata fixed effects. Lee bound samples are formed by truncating the sample of residualized scores after controlling for baseline covariates and strata controls. Imbens and Manski (2004) confidence intervals are given in brackets. Sample sizes for the Lee bounds estimates are the number of observations in the sample after truncating. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A5: Effects of scholarship and Bridge enrollment on fees and missed class (PP_{brig} & P_{brig} samples)

	PP_{brig} & P_{brig}			PP_{brig}			P_{brig}		
	Mean	ITT	2SLS	Mean	ITT	2SLS	Mean	ITT	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total edu expenditure (000s KES, pa)	11.349	-4.231***	-4.993***	10.910	-3.045***	-3.677***	11.489	-4.612***	-5.401***
		(0.403)	(0.473)		(0.786)	(0.949)		(0.469)	(0.544)
School fees (000s KES, pa)	6.413	-4.109***	-4.827***	6.582	-3.499***	-4.209***	6.359	-4.304***	-5.019***
		(0.287)	(0.337)		(0.545)	(0.659)		(0.336)	(0.391)
Missed class b/c unable to pay fees (pa)	0.340	-0.263***	-0.311***	0.370	-0.267***	-0.322***	0.331	-0.263***	-0.308***
		(0.022)	(0.027)		(0.047)	(0.056)		(0.026)	(0.030)
Missed class b/c unable to pay fees (pu)	0.106	-0.101***	-0.118***	0.107	-0.044	-0.052	0.106	-0.120***	-0.139***
		(0.015)	(0.018)		(0.032)	(0.038)		(0.017)	(0.020)
# days absent (past term, pu)	1.857	-0.799***	-0.934***	1.407	-0.167	-0.199	1.998	-0.988***	-1.149***
		(0.185)	(0.215)		(0.294)	(0.347)		(0.224)	(0.259)
>5 days absent past term (pu)	0.076	-0.043***	-0.050***	0.053	0.015	0.018	0.083	-0.060***	-0.070***
		(0.013)	(0.016)		(0.024)	(0.028)		(0.016)	(0.018)
>10 days absent past term (pu)	0.026	-0.031***	-0.036***	0.013	-0.013	-0.016	0.030	-0.036***	-0.042***
		(0.008)	(0.009)		(0.011)	(0.013)		(0.010)	(0.011)

Notes: All specifications control linearly for the randomization strata probability of scholarship assignment, and for planned 2016 grade dummies. The endogenous variable in 2SLS specifications is an indicator for *enrolling at Bridge for free*. Parent reported outcomes are indicated by (pa), pupil reported outcomes by (pu). Values for pupil absences that exceed 30 days are set to 30 days, approximately the 99th percentile. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A6: Effects of scholarship and Bridge enrollment on academic subject knowledge (P_{brig}, PP_{brig} & P_{priv} samples)

	P_{brig}			PP_{brig}			P_{priv}			
	EYS		SD	EYS		SD	EYS		SD	EYS
	ITT	2SLS	2SLS	ITT	2SLS	2SLS	ITT	2SLS	2SLS	gradient
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Math & language scores										
Subject knowledge index	0.10*	0.42*	0.38*	-0.07	-0.35	-0.32	0.15	0.57*	0.53*	0.92***
	(0.06)	(0.24)	(0.22)	(0.13)	(0.63)	(0.58)	(0.10)	(0.35)	(0.32)	(0.01)
Higher order skills index	0.09*	0.37*	0.41*	-0.04	-0.21	-0.23	0.15*	0.55*	0.61*	1.11***
	(0.05)	(0.20)	(0.23)	(0.12)	(0.57)	(0.64)	(0.08)	(0.31)	(0.34)	(0.01)
Language knowledge index	0.09	0.35	0.23	-0.21	-1.00	-0.65	0.16	0.62	0.40	0.65***
	(0.07)	(0.29)	(0.19)	(0.17)	(0.82)	(0.53)	(0.11)	(0.41)	(0.27)	(0.01)
Math score	0.11*	0.46*	0.49*	0.01	0.06	0.06	0.14	0.54	0.57	1.06***
	(0.06)	(0.26)	(0.27)	(0.14)	(0.64)	(0.68)	(0.10)	(0.35)	(0.37)	(0.01)
English score	0.12	0.48	0.28	-0.17	-0.82	-0.48	0.21*	0.80*	0.47*	0.59***
	(0.08)	(0.31)	(0.18)	(0.17)	(0.84)	(0.49)	(0.12)	(0.45)	(0.26)	(0.01)
Kiswahili score	0.06	0.23	0.14	-0.24	-1.18	-0.72	0.12	0.45	0.28	0.61***
	(0.09)	(0.36)	(0.22)	(0.18)	(0.86)	(0.53)	(0.12)	(0.44)	(0.27)	(0.01)
Observations		1,204			388			991		
Panel B: Science & social studies										
Science & social studies	0.08	0.34	0.18	-0.10	-0.62	-0.33	0.08	0.28	0.15	0.53***
	(0.11)	(0.45)	(0.24)	(0.06)	(0.43)	(0.23)	(0.12)	(0.43)	(0.23)	(0.01)
Science	0.05	0.20	0.17	-0.07	-0.45	-0.37	-0.00	-0.01	-0.01	0.83***
	(0.10)	(0.39)	(0.32)	(0.06)	(0.39)	(0.33)	(0.10)	(0.37)	(0.30)	(0.01)
Social studies	0.12	0.49	0.19	-0.10	-0.59	-0.23	0.13	0.47	0.18	0.39***
	(0.10)	(0.39)	(0.16)	(0.06)	(0.41)	(0.16)	(0.12)	(0.44)	(0.17)	(0.01)
Observations		1,203			205			973		
Local content score	0.07	0.30	0.12				0.13	0.53	0.22	0.41***
	(0.11)	(0.43)	(0.18)				(0.18)	(0.72)	(0.29)	(0.01)
Observations		1,030						610		

Notes: Test score outcomes are measured in standard deviations of the Standard 4 test score distribution. All specifications include controls for baseline characteristics, planned 2016 grade, and randomization strata fixed effect. The higher order skills index restricts excludes subject test items that reflect remembering/recalling information. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Appendix C Deviations from analysis plan

This appendix discusses deviations from the analysis plan described in the Pre-analysis Statement and Plan for Subsequent Analysis (PASPA).

C.1 Attrition

The PASPA was based on a sample that used an indicator for whether a pupil was included in the endline. This was used to identify the sample on which first stage and other preliminary analyses were calculated. After unblinding to the data, it was discovered that 59 observations that were indicated as being lost to follow-up did have test scores, although the interview was incomplete. We include these observations.

C.2 Strata controls

The PASPA specified that randomization strata indicators would be included in all analyses. However, in some cases, we control linearly for the probability of scholarship assignment in the randomization strata instead. In the analysis of the effects on academic achievement, this approach was adopted because strata dummy controls were causing some observations to drop out due to insufficient variation in the instrument and the precision of the estimate to decrease relative to linear probability controls. This is because, for many cohorts, the overall rate of secondary school transition and KCPE test taking is low which increases the probability that the instrument is perfectly predicted by the strata control.

C.3 Field officer controls

The PASPA discussed including field officer controls in some specifications because it was observed that field indicators were predictive of performance, especially on cognitive and non-cognitive assessments. However, after further consideration, the research team felt that field officer controls should not be included in primary specifications because the field officer that interacted with a pupil is potentially endogenous to treatment. The point estimates obtained for cognitive and non-cognitive outcomes are robust to including field officer controls. In the case of the head-knees

task, the results are not statistically significant when including the field officer controls, but are marginally significant when they are excluded.

C.4 Receptive Vocabulary

The analysis reports separate results by language (English and Kiswahili). This was not discussed in the the PASPA.

C.5 Alternative use tasks

We modified the scoring procedure for this task after completing the PASPA. Initially we had planned to score this task by counting the number of unusual uses as assessed by the field officer. However, inspection of the data suggested that protocol for grading this item were not uniformly applied. Field officers classified responses as either normal or unusual during the pupil interview. Some field officers scored tasks related to "eating" as unusual even though this was an example of a normal task provided during training, and field officers sometimes listed very similar uses twice despite the fact that training clearly stated that uses that involved similar motions should be counted only once. We clean the data to standardize scoring by automatically classifying responses concerning eating as "normal", and removing duplicate responses. Finally, we also provide an alternative score for this task that sums the total number of uses (a measure of ideational "fluency").

C.6 IRT estimation

In the PASPA we discussed results from estimation of IRT models, including equating coefficients. During replication of these results, we discovered that the R command `ltm` produced slightly different estimates than the corresponding command `irt` in Stata. Specifically, we identified large discrepancies in estimates of item discrimination terms, especially for items with high levels of discrimination. The Stata command appeared to systematically estimate higher discrimination for items identified as having high discrimination in R. Because R is more commonly used in the psychometrics literature, we now estimate the IRT model using R. Also simulations indicated that the Stata command may systematically overestimate discrimination terms for high discrimination items. We are deeply grateful to Peter Hickman for bringing this issue to our attention.

C.7 Quantile regression

The report contains results that estimate effects of attending Bridge on quantiles of the test score distribution using the quantile treatment effects method of Abadie et al. (2002). The PASPA indicated the procedure from Chernozhukov and Hansen (2013) would be used for this type of analysis. However, we determined that the Abadie et al. (2002) procedure was more appropriate in this context because it relies on the same assumptions used for identifying the local average treatment effect using 2SLS. The procedure from Chernozhukov and Hansen (2013) employs a different assumption of *rank similarity* that rules out strong correlations between individual treatment effects and take up (Wüthrich, 2020). Results using the approach from Chernozhukov and Hansen (2013) are also reported in the Online Appendix.

C.8 Pupil social attitudes

The PASPA indicated that we would evaluate the effect of attending Bridge on social attitudes. We present these results in the Online Appendix. Following the PASPA, we tested for an effect of the scholarship on pupil social attitudes. These items included measures of trust toward people in their own ethnic group, toward people in other ethnic groups, beliefs about gender disparities in intelligence, and support for democratic values. We found no evidence of effects on any of these outcomes.

C.9 Occupational aspirations

The PASPA indicated that we would evaluate the effect of attending Bridge on occupational aspirations. We present results on occupational aspirations in the Online Appendix. We find no evidence of effects on occupational aspirations.

C.10 Equivalent Years of Schooling

In the document, we report effects in Equivalent Years of Schooling (Evans and Yuan, 2019). The PASPA did not discuss this scaling of the effects. This choice scales the point estimates by a fixed factor and does not affect statistical inference.

C.11 Higher order skills index

The higher order skills index was developed after observing the effects on other test score outcomes and was not included in the PASPA. The classification of test score items used to construct the sub-test on which the higher order skills index is based was constructed before the PASPA was written.

C.12 Input indices

The PASPA specified that we would evaluate the effect of the scholarship on several outcomes related to inputs in education production. In the interest of brevity, we have combined many of these outcomes into indices.

The PASPA did pre-specify the construction of the adaptivity index. All other input indices (classroom crowding, teacher-classroom engagement, parental engagement, and test preparation) were constructed using the same standardization procedure as the instructional adaptation index after observing the results on individual items.

C.13 School enrollment variable construction

The process for constructing the endogenous school enrollment variables was slightly modified after writing the PASPA, which described the process in detail. The modifications produced changes in the classifications of eleven observations and identified data for seven observations that had been missing under the previous algorithm.

First, the script applied some steps only to pupils who were indicated as having been part of the endline survey. After unblinding to the data, it was discovered that some pupils recorded as not being in the endline were included in the endline survey. This change affected 3 observations.

Second, the algorithm used to determine attendance in 2017 used information on the type of school attended in 2016 combined with information at endline on whether the pupil had switched schools. If the pupil was attending a public or non-Bridge private school, the algorithm required that the 2017 phone call survey also indicate that the pupil did not switch schools. However, this rule was not applied when determining whether the pupil had attended Bridge. The research team chose to remove the restriction that the 2017 phone call survey corroborate the information. This

choice was made to ensure that that all school types were treated symmetrically and because the 2017 phone call survey had a relatively low follow-up rate meaning that a lack of corroboration generally coincided with missing data in the 2017 phone call survey.

C.14 Bridge school value-added

The PASPA described a plan quantify heterogeneity in academy-level effects using observational test score value-added estimates based on Bridge's administrative test score records. These results have been omitted because estimated value-added from years before the intervention were uncorrelated with analogous estimates using data from the time of the intervention, presenting serious challenges for interpretation, including the potential for attenuation bias due to measurement error.

C.15 Local education environment survey

The PASPA described plans to link results from a survey of other local education providers conducted in 2016. This analysis was not pursued due to challenges in matching school names and due to concerns regarding non-representativeness of the survey of local education providers.

C.16 Construction of the *main* sample

The PASPA contained an error in describing the construction of the *main* sample. The PASPA stated that the P_{main} sample was composed of any primary school aged pupil who said that they were planning to enroll in a public school, regardless of where they were enrolled at the time of the applicaton. The criteria used required that pupils be planning to enroll in public school *and* be currently enrolled in a public school.

C.17 Mean academy-intention counterfactual choices

This report uses only three academy-intention counterfactual means to form the instruments used to decompose the separate effect of Bridge and the payment of fees in section 7.2 and section 8, whereas the PASPA proposed the use of a fourth: the probability that a pupil was unenrolled. This variable was not included in the specifications because non-enrollment was very rare. In the P_{brig} and P_{priv} samples, no pupils were unenrolled in 2017, so that the empirical distribution of means

was degenerate and empirical Bayes shrinkage factor was undefined. Imputing the non-enrollment prediction to be zero, we find similar estimates as those from the three-variable specifications reported in this paper and the Online Appendix, although the F-statistics for the first stage tend to be smaller, as would be expected given the addition of an instrument with little marginal predictive power.