The Bridge Effect

A Comparison of Early Grade Learning Gains in English and Maths

Impact Evaluation Report

Lisa Chen, Lauren Nelson, and Stefanie Tung

The Bridge Effect

A Comparison of Early Grade Learning Gains in English and Maths

2013-14 Impact Evaluation Results

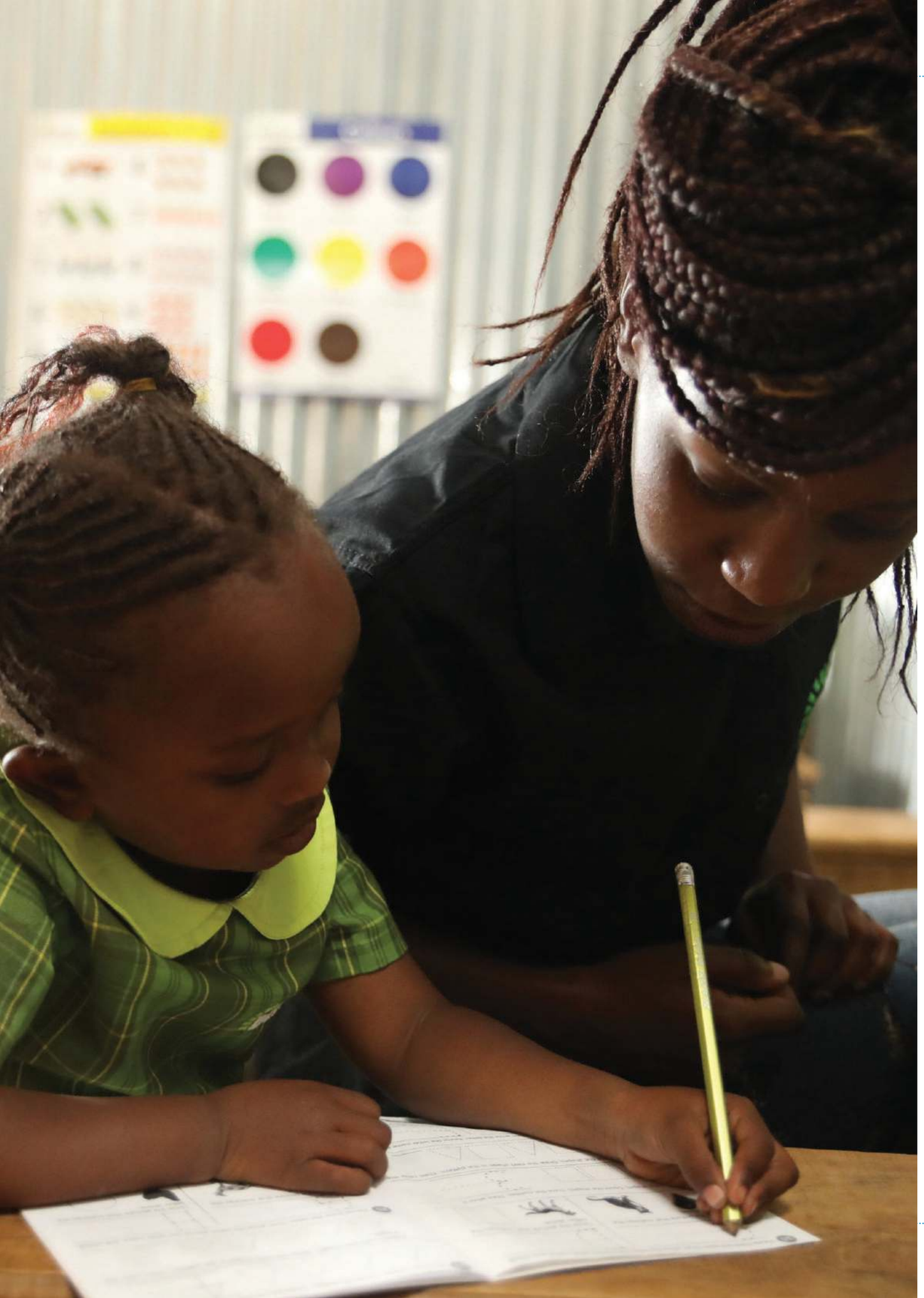
Lisa Chen, Lauren Nelson, and Stephanie Tung

November 2015

Acknowledgements: We would first like to thank neighbouring schools in the communities we serve for their participation in our EGRA/EGMA evaluation programme and their enthusiasm for using data to drive improvements in learning for their pupils. We thank Decisions Management Consulting for their work as an implementing partner to provide independent assessments of pupils, year after year; we are grateful for their leadership in the field. We also tremendously value the input and guidance of Andrew Fraker, Willa Friedman, Michael Kremer, and Mauricio Romero. Finally, this study would not have been possible without the contributions of Qatherine Dana, Beatrice Masters, and Tristan Zucker, who provided excellent research assistance over the course of the study. All viewpoints and any errors are our own.

Lisa Chen, Director, Measurement & Evaluation
Lauren Nelson, Senior Analyst, Measurement & Evaluation
Stephanie Tung, Manager, Measurement & Evaluation

Table of Contents



Executive Summary

The mission of Bridge International Academies is to provide affordable, high-quality education to millions of primary school pupils around the world. To hold ourselves accountable to our goals, we conduct one of the most comprehensive and rigorous early grade reading and math assessment programmes in East Africa. Using the Early Grade Reading Assessment and Early Grade Math Assessment (EGRA/EGMA), we measure the foundational literacy and numeracy skills of our pupils and those in neighbouring schools to understand both absolute and relative pupil performance. These tools were developed by the Research Triangle Institute (RTI) in conjunction with USAID and have been used by education ministries and multi-lateral agencies around the world.

Bridge's EGRA/EGMA programme includes a multi-year, pupil-matched study at Bridge International Academies and neighbouring public schools. Working in partnership with a third-party organization, each pupil is given a 30-minute, one-on-one oral assessment by an independent assessor. The programme then aims to follow the same pupils and measure their performance over time.

In this report, we highlight the results from the 2013-2014 administrations of EGRA/EGMA, based on a nationally representative sample of more than 2,700 pupils first assessed in October 2013. Because we could not randomly assign pupils to schools, we collected detailed demographic, education, and home life information in order to control for these factors. Given these factors and the pupils' incoming skill and ability levels, we find a Bridge effect of .31 standard deviations in English. This is equivalent to over 32% more schooling in one academic year. In maths, the Bridge effect is .09 standard deviations, or 13% additional schooling.

These effects are significant on both a statistical and practical level. They mean the average Bridge pupil experienced larger academic gains than their public school peers over the course of a year. Specifically, Bridge pupils learned to read almost 10 more words per minute and could also understand more of what they read, answering 10% more problems correctly than their public school counterparts. For maths, gains are focused on the quantity discrimination and word problems, where Bridge pupils answer 10.1% and 5.8% more questions correctly, respectively.

In addition, within one academic year, a greater proportion of Bridge pupils cross the English reading and maths fluency benchmarks set by key education officials in Kenya. On average, 30% and 31% of Bridge pupils in this study passed the fluent and emergent benchmarks for reading, compared to 16% and 24% of pupils in public schools, respectively. In maths, where benchmarks are based on addition and subtraction, the difference persists and ranges from 2.6 to 6.9 percentage points.

This study, spanning 13 counties in Kenya and adopting quasi-experimental methods to assess the efficacy of Bridge, represents one of the most rigorous studies conducted by any practitioner educational or development organization to date. We acknowledge that our results are not part of a randomised controlled trial, and that we face limitations due to high sample attrition rates. As such, we welcome constructive dialogue and engagement with stakeholders working to ensure that all pupils in Kenya are learning. We especially look forward to the next 5 years, during which we will be participating in a randomised evaluation in Kenya conducted by a team of leading external researchers.

I. Introduction

Bridge International Academies strives to provide affordable, high-quality education to children of households living on less than $2 a day. To hold ourselves accountable to this standard and to measure our efficacy, we have instituted an annual evaluation programme. Since 2010, Bridge has used the Early Grade Reading and Math Assessments (EGRA/EGMA) to measure pupils' foundational literacy and numeracy skills.

EGRA/EGMA reveals both absolute and relative pupil performance. Analysing absolute scores of Bridge pupils helps us determine the proportion of our pupils who are hitting literacy and numeracy benchmarks. Our EGRA/EGMA data also include scores of primary school pupils attending nearby public schools, allowing us to evaluate the relative performance of Bridge pupils with that of pupils from nearby schools. They allow us to answer the question: do our Bridge pupils perform better than they would have if they had attended a public school?

To answer this question, we launched an impact evaluation that gathered detailed information from individual pupils in 2013 and 2014. We improved the rigour and thoroughness of our EGRA/EGMA evaluation programme from previous years on three fronts – the sampling strategy, the execution of data collection, and the analysis of results. First, as Bridge grew to over 150 academies at the end of 2013, we updated our sampling strategy to account for geographic expansion and the need for additional comparison schools to detect meaningful effect sizes. Next, we implemented strict monitoring and quality control standards during the data collection period. Finally, we adopted the use of quasi-experimental analytical methods to measure our impact on pupil achievement, moving forward from a comparison of means.

In this report we detail the results of our 2013-2014 study. The report is organized as follows:

First, we discuss briefly the current state of primary education in Kenya and the complementary role of Bridge in improving educational quality. We then detail Bridge's on-going evaluation efforts, focusing on the 2013-14 EGRA/EGMA impact evaluation project. Next, we review this study in additional detail, including the instruments used, the sampling strategy, school participation, and attrition.

Section IV contains our analysis of pupil literacy and numeracy performance. We start by examining the pupils' baseline characteristics, then move onto describing our difference-in-differences model with the use of baseline control variables, and close with the results of the analysis.

Section V summarizes the results and offers different ways to think about the impact of Bridge on academic achievement, and Section VI examines these results from the perspective of meeting Kenyan education benchmarks as well as within the larger global context.

Given high sample attrition rates, we then examine differential attrition as a limitation to our study in Section VI. We also conduct robustness checks using a panel fixed effects model and Lee bounds. We conclude in Section VIII.

II. Background and Context

1. Primary Education in Kenya

Kenya established free primary education in 2003. Over a decade later, high levels of enrolment belie the fact that quality remains stubbornly low. The Uwezo initiative, which conducts household-based literacy and numeracy assessments across hundreds of communities in Kenya, has found that over two-thirds of children in Class 3 are not literate or numerate at the Class 2 level.[1] Even worse, after four-rounds of assessments from 2009 to 2013, Uwezo has found no improvement in the levels of learning for children in Kenya.

Fortunately, there are several efforts underway to improve the state of the primary education system in Kenya – from increasing pupil and teacher attendance to improving learning outcomes. Research institutions such as the Abdul Latif Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action (IPA) and the African Population Health and Research Centre (APHRC) evaluate strategies and provide evidence on what works. Organizations such as Evidence Action (EA) and the Research Triangle Institute (RTI) work in partnership with government ministries to scale up proven strategies, such as school-based deworming and new reading and maths initiatives.

Bridge International Academies strives to meet the need for evidence-based education. With 405 academies today across 44 counties in Kenya, we serve over 100,000 pupils in difficult-to-reach areas and are the largest network of complementary schools in the country. Our scale allows us to invest in critical educational infrastructure and in turn provide quality schooling at a cost affordable to parents. Our integration of technology with teacher and manager talent at our academies ensures that our pupils are learning. And at the core of Bridge, we strive for evidence-based decision making – both from our internal measurement efforts and by incorporating into our programming the findings and recommendations of these organizations.

2. The Bridge Approach

Bridge has developed systems and processes integrated with innovations in technology so that every pupil receives a high quality education. We believe a high quality education relies on dedicated teachers, engaging lessons, effective learning materials, a safe environment for learning, and ultimately, accountability to parents. Here, we examine each component of a high quality education and elaborate on how Bridge puts it into practice.

Dedicated Teachers

Mechanisms have been established throughout Bridge to ensure that there are teachers present in every classroom every day. Strict teacher attendance monitoring allows Headquarters staff to quickly identify academies where teachers are absent. A network of substitute teachers ensures that if the permanent teacher is not able to be at school, a substitute can be dispatched to cover their class. Furthermore, detailed lesson plans allow the substitute to continue where the permanent teacher left off.

1 Uwezo, Kenya (2013).

Engaging Lessons

Bridge brings new, engaging lessons aligned to the Kenyan 8-4-4 national syllabus to our classrooms on a daily basis. These lessons, designed by a team of Master Teachers and delivered wirelessly to classroom teachers, utilise strategies and activities that are proven to increase pupil understanding. For example, because poor automaticity is a barrier to understanding new concepts, we incorporate automaticity practice across all class levels in maths. For example, rapid automatic recall of multiplication tables ensures a smooth transition to more advanced maths topics such as operations using fractions.

Effective Learning Materials

Our materials, including textbooks, workbooks, homework books, and practical learning aides, encourage pupils to internalise and apply concepts to real life situations. In Early Childhood Education (ECE) classrooms, for example, pupils use flashcards to practice new vocabulary and benefit from engaging storybooks that introduce them to the joy of reading. Lower primary pupils use tens frames to develop a strong understanding of basic maths concepts, while upper primary pupils carry out experiments from the Bridge Science Kit.

Safe Environments for Learning

When pupils feel safe and supported in their learning environment, they are excited and engaged in school. At Bridge, corporal punishment is strictly forbidden. Our teachers and Academy Managers are equipped with a variety of behaviour management techniques that encourage pupils to correct their behaviour without any physical repercussions. Positive reinforcement of good behaviour sets the standard in the classroom and supports pupils to improve behaviourally as well as academically.

Accountability to Parents

Our parents have made the choice to send their child to Bridge, placing a great deal of trust in our Academy Managers and teachers. Our policy of no cash payments is one major way that we remain accountable to parents. Cashless payments protect parents by preventing fraud and ensuring that all the money they have paid goes directly towards training and supporting academy staff, creating effective lessons and learning materials, and maintaining the structure and grounds of the academy.

III. The 2013-14 EGRA-EGMA Impact Evaluation Project

As Bridge strives to provide quality education for all pupils, we constantly ask ourselves, “How do we know that the Bridge approach is working? How do we continuously learn and improve?”

Every day, we collect data from the ground up to understand the current academic performance of our pupils, drive curricular and training decision-making, and assess the effectiveness of our work. A key part of these efforts is our annual EGRA/EGMA programme, which includes an impact evaluation study. We work in partnership with a third-party organization, Decisions Management Consultants (DMC), to measure the foundational literacy and numeracy skills of our pupils and their academic peers in neighbouring schools. We then track the same pupils over time to measure learning gains – both within Bridge and relative to other schools in the same communities.

There are three cohorts of pupils in the programme: the first cohort was first assessed in July 2011 in the greater Nairobi area; the second cohort was first assessed in February 2012 in four major urban areas across Kenya (Nairobi, Nakuru, Eldoret, and Kisumu); and the third cohort was first assessed in October 2013 in several randomly selected counties across Kenya. This third cohort of pupils is representative of Bridge’s national presence and therefore serves as the sample for our impact evaluation study. Our most recent round of data collection in October 2014 aimed to survey these same pupils in order to measure their gains in one academic year. This paper focuses on the results of assessments conducted in 2013 and 2014, otherwise known as the 2013-14 EGRA/EGMA Impact Evaluation Project.

1. The Instrument

The survey instrument used in the impact evaluation project had three major components: 1) EGRA, 2) EGMA, and 3) Pupil Characteristics. Both EGRA and EGMA were developed by RTI in conjunction with USAID.[2] EGRA is an oral pupil assessment designed to measure the most basic foundational skills for literacy in the early grades: listening with comprehension, reading letters and simple words, and understanding sentences and paragraphs. EGMA is also an oral assessment designed to measure pupils’ foundational skills in numeracy and mathematics in the early grades, including quantity discrimination, addition, subtraction, and solving word problems. Each of these foundational skills is tested with a different component of the assessment, referred to as a “subtask.” The final component is a pupil survey collecting information on pupil demographics, education, and home life.

2 Research Triangle Institute (RTI) developed the EGRA/EGMA instruments and have applied them in 16 countries and in more than 20 languages. They have been used by other implementing partners in more than 40 other countries and more than 70 other languages.

a. Subtasks

Up to 5 EGRA subtasks and 6 EGMA subtasks were administered to pupils in classes Pre-Unit to Class 2.[3] See Appendix 2 for a detailed description of the subtasks administered in Bridge's EGRA/EGMA programme. The decision of which subtasks to administer to whom was made in consultation with members of the academic team. We took into consideration both the appropriateness of assessing subtasks at certain grade levels (would bottoming-out or ceiling effects occur?) and which set of subtasks would provide us with the most information on where our pupils stand. For example, because the subtasks have varying levels of difficulty and younger pupils tend to have shorter attention spans than older pupils, pupils in Pre-Unit in 2013 were administered fewer subtasks. See Table 1 for a list of the subtasks administered by grade.

Table 1. Subtask Administered in 2013-2014, by Class Level

Subtask	PU/C1	C1/C2	C2/C3
Listening Comprehension	X	X	
Non-Word Reading	X	X	
Familiar Word Reading	X	X	X
Passage Fluency	X	X	X
Reading Comprehension	X	X	X
Quantity Discrimination	X	X	
Addition Level I	X	X	X
Addition Level II		X	X
Subtraction Level I		X	X
Subtraction Level II		X	X
Word Problems		X	X
Total # of Subtasks	**7**	**11**	**9**

See Table 2 below for more details on the total number of questions and scoring criteria for each subtask. Some subtasks measure skills and knowledge, while others include a component of fluency by adding a time dimension to the subtask.

3 In Kenya, primary school class levels are formally known as Standard 1, 2, etc. but are commonly referred to as Class 1, 2, etc.

Table 2. Scoring Equations for Each Subtask Administered 4

Subtask	Scoring Formula	Total # of Questions
Listening Comprehension	(# Correctly Answered Questions)/(# of Questions Asked)	3
Non-Word Reading	(# Correct Words)/((60-remaining seconds)/60)	50
Familiar Word Reading	(# Correct Words)/((60-remaining seconds)/60)	50
Passage Fluency	(# Correct Words)/((60-remaining seconds)/60)	61
Reading Comprehension	(# Correctly Answered Questions)/(# of Questions Asked)	5
Quantity Discrimination5	(# Correctly Answered Questions)/(# of Questions Asked)	10
Addition I	(# Correctly Answered Questions)/((60-remaining seconds)/60)	20
Addition II	(# Correctly Answered Questions)/(# of Questions Asked)	5
Subtraction I	(# Correctly Answered Questions)/((60-remaining seconds)/60)	20
Subtraction II	(# Correctly Answered Questions)/(# of Questions Asked)	5
Word Problems	(# Correctly Answered Questions)/(# of Questions Asked)	5

b. Pupil Characteristics

In 2013, we began collecting more detailed background information on pupils participating in the study.[6] We asked questions about pupil demographics (gender, age), education (ECE attendance, grade level), household characteristics (ownership of a radio or television, whether the pupil had electricity), activities at home (listening to the radio, watching television, reading, doing homework, and with whom), and meals consumed within a day of the assessment (number of meals and types of foods consumed).[7]

We collected this information to allow later analysis by sub-groups within the sample. These observables provide clarity on any systematic differences between the groups receiving and not receiving the Bridge "treatment" that may influence results.

4 Timed subtasks allot 60 seconds.

5 Although the Quantity Discrimination subtask used in 2013 has 10 questions, the scoring used for this analysis incorporates pupil responses to the first 7 questions only. Therefore, each pupil's score on Quantity Discrimination was scored as (# of questions correct) / 7. This decision was made due to a Tangerine assessment rendering error in which the last 3 questions of Quantity Discrimination were truncated from our electronic assessment tool. To correct for this, the enumerators were instructed to manually input each pupil's response to questions 8, 9, and 10 into a comments box at the end of each assessment. Because some completed assessments did not have the pupil responses entered, however, the responses to these questions were not incorporated into the scoring for this subtask for any pupils.

6 Previous EGRA/EGMA administration rounds limited data collection to school names, locations, and pupil assessments.

7 Demographic questions were included at the beginning and end of the survey instrument. We selected these questions based on RTI's instrument, consultations with researchers, and piloting on Bridge pupils. The questions included could all be reasonably and accurately answered by young pupils.

2. Sampling Strategy and Randomization

In anticipation of the 2013 round of data collection, we conducted power calculations to determine an adequate sample size for obtaining new baseline data. We set our power at 80% and confidence level at 5%, which is the standard for impact evaluations in practice. While the assessments and outcomes are on the pupil level, the Bridge "treatment" is at the school level, and therefore we took into consideration clustering. We found that, for a given sample size, we needed a greater number of schools as opposed to a greater number of pupils at any given school. Also, as expected, the smaller the effect size we sought to detect for any given subtask, the greater the sample size requirement. Acknowledging operational and resource constraints, we decided upon a sample of 4,536 pupils across 84 schools.

a. Selection of Schools

To maximize power for our sample size, we wanted an equal number of schools in each comparison group: 42 Bridge International Academies and 42 public schools.

From here, we did our best to minimize bias in our sampling procedure, given operational realities and incomplete information. Because regional differences were still quite prevalent at the level of former provinces, we aimed for representation in all provinces where Bridge had a presence. We then randomly selected 1-2 counties within each former province from which to collect data.[8] Within counties, we randomly selected 3-4 Bridge International Academies (with a balance between new and existing academies) and corresponding comparison schools.

Table 3 shows the counts of schools and counties in our randomization, by province.

Table 4 shows the name of the counties randomly sampled, and the count of Bridge International Academies we sampled within each county.

Table 3. Number of Counties and Schools to Randomly Select

Former Province	# of Counties	Total # of Schools	# Bridge	# Public
Rift Valley	2	14	7	7
Nyanza	2	14	7	7
Central	2	12	6	6
Nairobi	1	14	7	7
Western	2	12	6	6
Coast	2	12	6	6
Eastern	1	6	3	3
Total	**12**	**84**	**42**	**42**

8 We only selected one county for the former Nairobi and Eastern provinces. Nairobi became its own county and Bridge was only present in one county of the former Eastern province at the time.

Table 4. Number of Bridge International Academies Sampled, by County and Province

Former Province	County	# Bridge
Central	Kiambu	3
Central	Nyandarua	3
Coast	Mombasa	3
Coast	Taita-Taveta	3
Eastern	Meru	3
Nairobi	Nairobi	7
Nyanza	Kisii	4
Nyanza	Nyamira	3
Rift Valley	Kajiado	3
Rift Valley	Uasin Gishu	4
Western	Bungoma	3
Western	Kakamega	3
Total		**42**

Our next step was to identify neighbouring public schools to serve as a comparison group. Because there is not a reliable, publicly available dataset that shows every school operational in Kenya, we used the research of our internal Customer Insights department, which maps the educational landscape in the communities we serve. For every selected Bridge International Academy, we generated a randomly ranked list of public schools to invite to participate in the study.

We then turned over the list to Decisions Management Consulting (DMC), a third-party organization specializing in education and fieldwork in Kenya. DMC was tasked with securing the appropriate national and local permissions to conduct assessments at comparison schools.

Once in our communities, DMC project managers approached the headmaster of the schools in the random order of the list until a school was willing and able to participate.[9] If a list was exhausted with no acceptances, DMC was asked to locate a suitable comparison school due to their field experience. DMC could add additional suitable comparison schools should there be too few pupils in the existing set of participating schools.

In return for taking part in our study, each comparison school would receive a detailed report on its pupils' performance, complete with comparative statistics on performance of other schools within the same county. All school names in both these reports and this study are kept anonymous to maintain data integrity and encourage participation in future rounds of assessments.

9 Some schools declined to participate. Others were ineligible due to their existing participation in PRIMR and/or if they did not have a sufficient number of pupils.

b. Selection of Pupils within a School

Our target was for DMC to assess 54 randomly selected pupils at each school: 18 pupils per grade level Pre-Unit through Class 2 in 2013, stratified by gender. The on-the-spot randomization process for selecting pupils is detailed in Appendix 1. The assessors (and not the school teachers) were tasked with selecting the pupils to prevent biased selection of pupils within the school. Since schools had differing class sizes, we allowed DMC to oversample pupils at another school of the same school type within the same county if needed to make up for shortfalls.

3. School Participation and Pupil Tracking

DMC was charged with training the assessors in the instrument, coordinating fieldwork, and providing oversight of field teams.[10] To streamline the data collection process, assessors were provided with tablets and used the Tangerine[11] platform to conduct surveys and assess pupils. Bridge's internal M&E team established a number of processes to ensure data integrity by monitoring DMC's daily progress. These efforts were enabled by electronic data collection. DMC uploaded completed assessments daily so the M&E team could react to data in real time and communicate with DMC to address issues in 1 to 3 days. For more details, see Appendix 3.2.

a. Participating Schools

A total of 88 Bridge and public schools participated in the 2013 EGRA/EGMA assessment round and 87 of these schools continued their participation in 2014. See Table 5 below.

Table 5. Count of Participating Schools by Province, School Type, and Year[12]

		Participating in 2013			Participating in 2013 & 2014		
Former Province	**# of Counties**	**# BIA**	**# Public**	**Total # Schools**	**# BIA**	**# Public**	**Total # Schools**
Rift Valley*	3	10	3	13	10	3	13
Nyanza	2	7	6	13	7	6	13
Central	2	6	4	10	6	3	9
Nairobi	1	11	8	19	11	8	19
Western	2	6	8	14	6	8	14
Coast	3	6	7	13	6	7	13
Eastern	1	3	3	6	3	3	6
Total	**14**	**49**	**39**	**88**	**49**	**38**	**87**

10 This included confirming that schools were visited and that the pupils were assessed. It also included ensuring that the data were saved and uploaded on a daily basis.

11 Developed by RTI, Tangerine is a software application that allows the programming of tailored EGRA/EGMA instruments online, which then can be downloaded onto tablets and used to collect field data in real time. http://www.tangerinecentral.org/

12 Two out of three counties from Rift Valley had participating public schools. See Table 6 for more information.

Table 6 breaks down these counts further, by county. Between 2013 and 2014, one school declined to participate in Kiambu County. For the purposes of this report, we will only analyse data from the 87 schools that participated in both years.

Table 6. Count of Participating Schools by County[13, 14]

		Participating in 2013			Participating in 2014		
Former Province	**County**	**BIAs**	**Public**	**Total**	**BIAs**	**Public**	**Total**
Central	Kiambu	4	2	6	4	1	5
Central	Nyandarua	2	2	4	2	2	4
Coast	Mombasa	2	2	4	2	2	4
Coast	Kilifi*	1	1	2	1	1	2
Coast	Taita-Taveta	3	4	7	3	4	7
Eastern	Meru	3	3	6	3	3	6
Nairobi	Nairobi	11	8	19	11	8	19
Nyanza	Kisii	4	5	9	4	5	9
Nyanza	Nyamira	3	1	4	3	1	4
Rift Valley	Kajiado	3	1	4	3	1	4
Rift Valley	Uasin Gishu	6	2	8	6	2	8
Rift Valley	Laikipia*, +	1	0	1	1	0	1
Western	Bungoma	3	3	6	3	3	6
Western	Kakamega	3	5	8	3	5	8
Total	**14 Counties**	**49**	**39**	**88**	**49**	**38**	**87**

13 For schools located along the border of county boundaries, there were some inconsistencies in the reported county that the school operated under. This was understandable in the wake of the devolution which occurred in 2013. As such, we have since clarified county associations, but it unintentionally added participating schools in Laikipia and Kilifi, which were not in our original set of sampled counties.

14 There were no participating public schools in Laikipia, as this county was not originally sampled to be part of the impact evaluation. The Bridge academy in this county became part of this study prior to clarification of borders post-devolution.

b. Pupils Assessed in 2013-14

DMC successfully assessed 2,737 pupils in 2013 from the 87 schools participating in both years.[15] In preparation for the 2014 round of assessments, Bridge provided DMC with a list of all pupils assessed in 2013, which included detailed pupil-level information. Before the assessments began, assessors asked a series of questions to both the headmaster and pupil to verify the identity of the pupil using this information. The assessor would then determine whether the pupil was indeed the same child that had been previously assessed in 2013.

Unfortunately, we faced high sample attrition rates given the highly mobile population we serve coupled with school participation constraints and pupil absenteeism. Overall, about 41% of our sample attrited between 2013 and 2014. Because it was cost-prohibitive to track pupils outside the schools from which they were first assessed, we could not conduct follow-up assessments on pupils who had moved schools. In addition, DMC was also unable to assess pupils absent on the day of the visit, due to field operations and transportation budget constraints. Table 7 shows the count of pupils assessed in 2013 along with count of pupils who stayed versus left the sample by 2014. Table 8 shows the attrition rates by school type and class level.

Table 7. Count of Pupils Assessed in 2013 and Attrited in 2014

School Type	All Pupils Assessed in 2013				Assessed in 2013-14				Attrited Between 2013-14			
	PU	C1	C2	All	PU	C1	C2	All	PU	C1	C2	All
Bridge	757	795	838	2390	391	427	470	1288	366	368	368	1102
Public	699	757	804	2260	438	505	506	1449	261	252	298	811
Total	**1456**	**1552**	**1642**	**4650**	**829**	**932**	**976**	**2737**	**627**	**620**	**666**	**1913**

Table 8. Sample Attrition Rates from 2013-14[16]

School Type	Pupil-Level Attrition Rate				Average School-Level Attrition Rate			
	PU	C1	C2	All	PU	C1	C2	All
Bridge	48.3%	46.3%	43.9%	46.1%	47.8%	46.0%	43.3%	46.3%
Public	37.3%	33.3%	37.1%	35.9%	36.8%	31.9%	31.6%	34.5%
Average	**43.1%**	**39.9%**	**40.6%**	**41.1%**	**42.9%**	**39.8%**	**38.3%**	**41.1%**

The attrition rates at Bridge were higher than those at public schools and warrant further investigation. More discussion is in Section VII.

15 The original sample size goal was 6,804, which DMC met if 5 of the original participating schools had not declined to participate in 2014. This means that we are less likely to be able to detect effects of subtasks with higher intracluster correlations, which require a larger sample size.

16 The pupil-level attrition rate is defined as the count of pupils who were assessed in 2013 minus the count of pupils assessed in both 2013 and 2014, all divided by the count of pupils who were assessed in 2013. This is the pure sample attrition rate. The average school-level attrition rate follows the same logic but is calculated first at the school level, then averaged across all schools of that school type and grade. This is the average rate of attrition for any individual school in a single year.

IV. Analysis of Panel Data

The panel analysis limits our comparison to pupils who were assessed in *both* 2013 and 2014 at the 87 schools in the sample. The panel data includes 2,737 pupils from 87 schools.[17] These pupils make up the dataset used to conduct our panel analyses.

1. Summary Statistics

Bridge pupils are systematically different from pupils attending public schools, though the differences vary across characteristic. For more details on baseline characteristics, see Appendix A4.

- **Age:** Non-attriting pupils attending public schools are older on average than those attending Bridge, again suggesting a higher initial age of entry into the formal education system.
- **ECE:** A greater percentage of non-attriting pupils at public schools have attended at least one year of ECE. Again, over 90% of pupils in our communities have received at least some ECE.
- **Health:** Pupils across both school types consumed an average of 3 meals within a 24-hour period of the EGRA/EGMA assessment. Bridge pupils were slightly more likely to have had breakfast compared to public school pupils.
- **Language:** Pupils at Bridge were more likely to have English language exposure, and relative to public school pupils more likely to speak English at home.
- **Reading & Homework:** Conditional on receiving homework help or being read to at home, Bridge pupils are more likely to receive support from a parent.
- **Other Home Activities:** There is no clear pattern of systematic differences in other home activities with parental participation.
- **Assets and Electricity:** A greater proportion of Bridge pupils own televisions or have electricity than public school pupils.

These figures continue to suggest that Bridge pupils may start "better off" than their peers at public schools along these dimensions.

Moving to baseline test scores in English, Bridge pupils performed higher on average at baseline than their public school counterparts. Bridge pupils also start out higher at baseline in maths in comparison to public school pupils.

Baseline test scores by school type for non-attriting pupils are shown in Appendix 4.2, Table 23.

17 This includes 829 pupils who started Pre-Unit, 932 who started Class 1, and 976 who started in Class 2.

2. Panel Difference-in-Differences

a. Concept

Difference-in-Differences

The difference-in-differences (DiD) model allows us to estimate the "Bridge effect" – the change in pupil EGRA/EGMA performance as a direct result of attending Bridge – without having conducted a randomised experiment.

A DiD model isolates treatment effects by comparing the difference in outcome measures at two points in time for the treatment and control groups against one another. In other words, a DiD model takes the outcome measure at the first point in time for each group and subtracts it from the outcome measure at the second point in time for the same group. It then compares the within-group difference across the two groups to parse out the effect of the treatment. Average treatment effects are calculated as follows:

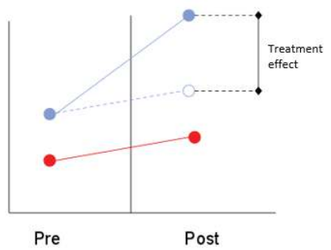
$$\textit{treatment effect} = (T_{t2} - T_{t1}) - (C_{t2} - C_{t1})$$

, where

T_{t2} = treatment group outcome at second point in time;
T_{t1} = treatment group outcome at first point in time;
C_{t2} = control group outcome at second point in time; and
C_{t1} = control group outcome at first point in time.

The estimation of treatment effects using a DiD model also can be depicted graphically, as shown in Figure 1 below. The red points are the control and the blue are the treatment:

Figure 1. Estimating Treatment Effects Using DiD Approach



DiD models ensure that any unobserved variables that remain constant over time and that are associated with the outcome will not bias estimates of treatment effects. Simple differences in within-group differences or a regression framework. As compared to calculating the simple difference in within-group differences, a regression framework provides us with the added benefit of including extra explanatory (control) variables in our model. Without control variables added into the framework, the estimation of treatment effects using either method is the same.

doing so, the model assumes that these unobserved variables would affect both the treatment and control groups equally. In other words, it assumes that if the treatment group had not received the treatment at all, its outcome would be the same as the outcome for the control group.[18]

By construction, all DiD models require data for at least two points in time. Without it, we cannot calculate the within-group difference, which relies on data captured at the first and second points in time. There are two major types of data that can serve this function: (1) repeated cross-sectional data and (2) panel data. For the analyses presented in this report, we chose to implement the one which draws upon panel data. A description of each of our selected model variations is described below.

Panel Difference-in-Differences

In a repeated cross-sectional DiD estimation, individuals do not need to be tested at both time periods. In a panel DiD model, however, we incorporate data at two time points from the exact same individuals. This type of model naturally excludes pupils who attrited from our sample as well as those who had not been tested previously. Therefore, the overall number of pupils included in our panel DiD model is smaller than the number of pupils originally tested in 2013.

There are multiple benefits to using a panel DiD approach for estimating treatment effects. A panel DiD model is both more straightforward to construct and more statistically powerful than a repeated cross-sectional DiD model.[19]

18 Buckley, Jack & Yi Shang (2003). "Estimating policy and program effects with observational data: the 'differences-in-differences' estimator." Practical Assessment, Research & Evaluation, 8(24). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=24.

19 Buckley, J. & Shang, Y. (2003)

b. Specification

In general, the linear equation for estimating DiD using panel data is:

$$outcome_{t2} - outcome_{t1} = \beta_0 + \beta_1(treatment) + \varepsilon$$

, where

$outcome_{t2}$ = outcome measure for all units at second point in time;
$outcome_{t1}$ = outcome measure for all units at first point in time; and
β_1 = effect of treatment.

The outcome measure is essentially the change in pupil performance (gains) by subtask.

We extend this model to control for additional factors that may influence the reported treatment effects. Aside from the benefits of implementing a panel DiD approach more generally, this particular extension of a panel DiD model is desirable because it takes into account additional variables that we suspect may have an impact on the treatment effect. It also allows us to factor in the possibility of differential trends depending on the pupils' characteristics at baseline. For example, it helps us understand whether pupils who start with lower baseline scores achieve an additional increase in performance at follow-up when compared to pupils as a whole.

The linear equation for estimating DiD using panel data with additional covariates is therefore:

$$outcome_{t=2} - outcome_{t=1} = \beta_0 + \beta_1(treatment) + \beta_2(covariate) + \ldots + \beta_x(covariate) + \varepsilon$$

, where

$outcome_{t=2}$ = outcome measure for all units at second point in time;
$outcome_{t=1}$ = outcome measure for all units at first point in time;
β_1 = effect of treatment;
β_2 = effect of first covariate on the difference in outcomes; and
β_x = effect of x^{th} covariate on the difference in outcomes.

The outcome variable at time t=1 is essentially a covariate and can be moved to the other side of the equation, and the linear equation now becomes:

$$outcome_{t=2} = \beta_0 + \beta_1(treatment) + \beta_2(outcome_{t=1}) + \ldots + \beta_x(covariate) + \varepsilon$$

Our resulting panel DiD estimator is based on the above linear equation, controlling for the baseline test score along with its squared and cubed functions, in order to allow for the possibility of non-linear effects. We also include other pupil characteristics collected at baseline, such as demographics, household assets, education, and health. The specification is as follows:

$$endline\ subtask\ score = \beta_1(treatment) + \beta_2(baseline\ subtask\ score) + \beta_3(baseline\ subtask\ score^2) + \beta_4(baseline\ subtask\ score^3) + \beta_4(class\ 1) + \beta_5(class\ 2) + \beta_6(female) + \beta_7(\#\ years\ of\ ECE) + \beta_8(has\ radio) + \beta_9(has\ TV) + \beta_{10}(has\ electricity) + \beta_{11}(\#\ meals) + \beta_{12}(DDS) + \varepsilon$$

Again, the treatment effect that we seek to isolate is the "Bridge effect". In this estimation, the coefficient β_1 representing the "Bridge effect" accounts for pupils' baseline score along with other pupil characteristics. Therefore, it is interpreted as the marginal effect of attending Bridge on gains in pupil EGRA/EGMA performance, holding each pupil's baseline score and other characteristics constant.[20]

c. Results

We analysed the output generated from our panel DiD model in order to compare Bridge and public school differences. Across seven out of eleven subtasks, the "Bridge effect" is large, significant, and positive at the 5% level. Among EGRA subtasks, the two most notable effects are in listening comprehension and reading comprehension – on average, Bridge pupils experienced an increase in performance of about .49 and .39 standard deviations in relation to their public school counterparts, respectively. Effects on non-word reading, familiar word reading, and passage fluency are also large and significant at .15, .24, and .31 standard deviations, respectively. For EGMA, there are large effects for quantity discrimination at .32 standard deviations and word problems at .20 standard deviations.

As for pupil-level characteristics, there are few clear trends. The inclusion of baseline test scores likely absorbs much of the information these additional observables provide. While it is easy to see that older pupils tend to perform better at maths, being female is associated with higher scores only for familiar word reading. Asset ownership is only positively associated with performance in listening comprehension, reading comprehension, and word problems. See Table 9 below.

20 As with the repeated cross-sectional method, we use STATA's -reg- command to run our model, clustering at the school level and reporting robust standard errors.

Table 9. Panel DiD, Bridge vs. Public Schools

	(1) Listening Comp.	(2) Non-word Reading	(3) Word Reading	(4) Passage Fluency	(5) Reading Comp.	(6) Quantity Disc.	(7) Add Level 1	(8) Add Level 2	(9) Sub Level 1	(10) Sub Level 2	(11) Word Problems
Bridge	**0.489****	**0.147****	**0.237****	**0.305****	**0.394****	**0.322****	**0.0763**	**0.0501**	**0.139+**	**0.117**	**0.198***
	(0.0923)	**(0.0479)**	**(0.0356)**	**(0.0411)**	**(0.0550)**	**(0.0558)**	**(0.0652)**	**(0.0707)**	**(0.0767)**	**(0.0857)**	**(0.0794)**
Baseline	0.420**	0.170**	0.110**	0.210**	0.472**	0.240**	0.117*	--	--	--	--
Task Score	(0.0652)	(0.0474)	(0.0352)	(0.0472)	(0.0528)	(0.0568)	(0.0524)	--	--	--	--
Baseline	--	--	0.0633	0.187**	0.564**	--	0.287**	0.304**	0.230**	0.364**	0.182**
Task Score ^2	--	--	(0.0541)	(0.0543)	(0.0646)	--	(0.0780)	(0.0526)	(0.0573)	(0.0563)	(0.0640)
Baseline	0.221**	0.812**	0.829**	0.889**	0.440**	0.258**	0.482**	0.168*	0.317**	0.152*	0.264**
Task Score ^3	(0.0680)	(0.0640)	(0.0395)	(0.0270)	(0.0361)	(0.0375)	(0.0606)	(0.0694)	(0.0638)	(0.0650)	(0.0517)
Pupil in Class 1	-0.0748	-0.189**	-0.249**	-0.147**	-0.169**	-0.0692+	-0.0540*	-0.0716*	-0.0546*	-0.0539	0.00706
	(0.110)	(0.0448)	(0.0368)	(0.0326)	(0.0522)	(0.0397)	(0.0206)	(0.0290)	(0.0263)	(0.0366)	(0.0315)
Pupil in Class 2	0.00361	0.00735	0.0224	0.00646*	0.0353	-0.0262	0.0110	0.0168	0.00559	0.0111	-0.00382
	(0.0586)	(0.0346)	(0.0220)	(0.00271)	(0.0227)	(0.0227)	(0.0195)	(0.0381)	(0.0223)	(0.0341)	(0.0244)
Age	-0.0207	-0.0287	0.00260	0.00860	-0.0171	0.0148	0.0270*	0.0512*	0.0648*	0.0587**	0.0480+
	(0.0186)	(0.0174)	(0.0102)	(0.0162)	(0.0176)	(0.0174)	(0.0132)	(0.0209)	(0.0275)	(0.0219)	(0.0251)
Female	-0.0355	0.0129	0.0535*	0.0410	-0.0194	-0.0203	-0.00487	-0.0142	-0.0601	-0.00609	-0.0314
	(0.0461)	(0.0419)	(0.0219)	(0.0335)	(0.0338)	(0.0349)	(0.0349)	(0.0449)	(0.0444)	(0.0422)	(0.0484)
Years of ECE	0.0308	-0.00931	0.000129	-0.0192	0.0305	0.0324+	0.00135	0.0127	-0.0105	-0.00115	0.0342
	(0.0275)	(0.0205)	(0.0138)	(0.0237)	(0.0210)	(0.0191)	(0.0178)	(0.0266)	(0.0293)	(0.0262)	(0.0293)
Has Radio	0.00102	-0.0337	-0.0458*	-0.0340	-0.0180	0.00437	0.00915	0.0207	-0.0799	0.0136	-0.113+
	(0.0500)	(0.0418)	(0.0222)	(0.0306)	(0.0395)	(0.0367)	(0.0356)	(0.0516)	(0.0803)	(0.0634)	(0.0615)
Has TV	0.133*	0.0582	0.0150	0.0447	0.0999*	-0.0236	0.0537	-0.0526	0.0716	0.0729	0.177*
	(0.0628)	(0.0431)	(0.0287)	(0.0587)	(0.0433)	(0.0401)	(0.0350)	(0.0558)	(0.0538)	(0.0617)	(0.0685)
Has Electricity	0.125+	-0.00674	0.0618+	0.0173	0.0892*	0.0776+	0.0884*	0.0896	0.0940	0.0364	0.0169
	(0.0684)	(0.0404)	(0.0335)	(0.0584)	(0.0393)	(0.0449)	(0.0388)	(0.0589)	(0.0678)	(0.0661)	(0.0684)
Meal Count	0.0599	-0.0130	-0.0113	0.0340	0.0256	0.0316	0.101*	0.0632	0.0714	0.161*	0.123
	(0.0767)	(0.0690)	(0.0431)	(0.0429)	(0.0566)	(0.0518)	(0.0479)	(0.0838)	(0.0690)	(0.0782)	(0.0763)
Diet Diversity Score	0.0177	-0.00209	0.0120	-0.00728	0.0413+	0.0275	-0.0287	-0.0336	-0.0834	-0.0512	-0.0417
	(0.0360)	(0.0250)	(0.0187)	(0.0178)	(0.0248)	(0.0226)	(0.0261)	(0.0322)	(0.0534)	(0.0394)	(0.0395)
Constant	-0.351	0.732**	0.444**	0.247	-0.276	-0.233	-0.303+	-0.378	-0.384	-0.759*	-0.654*
	(0.278)	(0.238)	(0.148)	(0.175)	(0.224)	(0.194)	(0.169)	(0.292)	(0.245)	(0.346)	(0.292)
Count	1672	1672	2626	2626	2329	1672	2626	1849	1849	1848	1848
R-squared	0.171	0.442	0.623	0.479	0.344	0.283	0.289	0.136	0.119	0.108	0.122

*Note: Standard errors in parentheses; + p<0.10, * p<0.05, ** p<0.01. Omitted class level dummy is Pre-Unit as of 2013, except for regressions 8-11, which omit the Class 1 dummy as of 2013 since Pre-Unit pupils were not administered these subtasks.*

V. Impact of Bridge on Academic Achievement

1. Summary of Effect Sizes

Table 10 below summarizes the Bridge effect size (the coefficient of the Bridge treatment variable) for each subtask and the average effect for EGRA and EGMA. Effects that are not statistically significant at the p < 0.05 level are shown as zeros and are used as zeros in calculating average effects.

Table 10. Bridge Effect Sizes by Subtask[21]

Subtask	Effect Size
Listening Comprehension	0.49
Non-word Reading	0.15
Familiar Word Reading	0.24
Passage Fluency	0.31
Reading Comprehension	0.39
Quantity Discrimination	0.32
Addition Level 1	0
Addition Level 2	0
Subtraction Level 1	0
Subtraction Level 2	0
Word Problems	0.20
EGRA Effect Size, p<.05	**0.31**
EGMA Effect Size, p<.05	**0.09**

We are aware that a statistically significant effect size does not necessarily equal causal impact. The only time they are truly equal is with a well-designed and well-implemented randomised controlled trial (RCT). That said, we believe the scale of our data collection efforts combined with the rigour of our analysis significantly moves us toward identifying the Bridge effect. With a large sample of pupils successfully tracked over the period of one year, we can more rigorously measure the impact of Bridge by holding the pupils' characteristics constant over time. This reduces the possibility of bias, for example, if different types of pupils attrite in 2013 than come in in 2014.

21 All subtasks without a statistically significant effect size at the 5% level are noted here as zeros.

2. Translation of Effect Sizes

While effect sizes are useful in framing the "Bridge effect" in the context of other educational interventions, they are intuitively more difficult to understand. Therefore, in the following two sections, we translate our results from effect sizes into units of measurement that are more familiar to a broader audience: absolute gains, additional days of learning, and percent of additional schooling.[22]

a. Gains in Literacy and Numeracy

For subtasks where there is a significant difference in gains between Bridge pupils and public school pupils, we translate the effect size back into non-standardized units. Our pupils have greater English reading fluency. When reading words in isolation, they read at least 5 more words per minute than their peers attending public schools. When reading text in a passage, they read almost 10 more words correctly per minute. Furthermore, they understand more of what they read. They are able to answer over 10% more questions correctly after reading a passage.

In maths, they are better able to distinguish the magnitude of numbers and answer 10% more questions correctly. Our pupils perform better on word problems, answering 5.8% more questions correctly. For other maths subtasks, there is no difference in the performance of pupils attending Bridge or public schools.

Table 11. Gains by Subtask, Bridge vs. Public Schools

Subtask	Gains
Listening Comprehension	14.6% more problems correct
Non-word Reading	2.1 more words correct per minute
Familiar Word Reading	5.6 more words correct per minute
Passage Fluency	9.6 more words correct per minute
Reading Comprehension	10.3% more problems correct
Quantity Discrimination	10.1% more problems correct
Addition Level 1	N/A
Addition Level 2	N/A
Subtraction Level 1	N/A
Subtraction Level 2	N/A
Word Problems	5.8% more problems correct

22 See Appendix A6. Guide to Effect Sizes for an in depth understanding of the reasoning for this process and the exact methods of calculating gains.

b. Additional % of Schooling

We can also express gains as a percentage of additional schooling (or growth), using the annual gains for public school pupils (shown in Table 12). For English, the Bridge effect of .31 standard deviations is equivalent to 32.3% additional schooling (or pupil growth) for lower primary pupils. For maths, the effect size of .09 standard deviations translates to approximately 13.1% additional schooling.

Table 12. Annual Gains For Public School Pupils Based on EGRA/EGMA Subtasks

	Weighted Average[23]	PU-C1	C1-C2	C2-C3
Listening Comprehension	0.90	0.81	0.98	N/A
Non-word Reading	0.88	1.03	0.75	N/A
Familiar Word Reading	1.10	2.11	0.93	0.41
Passage Fluency	1.17	1.86	1.15	0.60
Reading Comprehension	0.82	1.16	0.98	0.49
Quantity Discrimination	0.81	1.07	0.59	N/A
Addition Level I	0.75	1.00	0.66	0.63
Addition Level II	0.71	N/A	0.80	0.62
Subtraction Level I	0.53	N/A	0.51	0.55
Subtraction Level II	0.69	N/A	0.66	0.71
Word Problems	0.48	N/A	0.56	0.40
EGRA	**0.98**	**1.39**	**0.96**	**0.50**
EGMA	**0.66**	**1.03**	**0.63**	**0.58**

c. Additional Days of Learning

Another way to translate effect sizes into a more assessable figure is to convert the percent of additional schooling into days of learning. With approximately 200 school days in the academic calendar year in Kenya, the Bridge effect for lower primary translates to 64 and 26 additional days of learning in English and maths, respectively.

23 Weighted average calculated for each subtask by multiplying the number of pupils in a class by the class's average score, summing these values, then dividing by the total number of pupils taking the subtask in all classes.

VI. Comparative Results on Performance

1. Benchmarks in Kenya

In August 2012, RTI and the Kenya Ministry of Education (MoE) formed a panel comprised of Kenya National Examinations Council assessment experts, donors, and non-government organisation (NGO) community members. Led by the MoE and the Directors of Basic Education and Quality Assurance and Standards, the panel set Class 2 reading standards in English and Kiswahili.[24] EGRA benchmarks were determined based on findings from the PRIMR baseline study, particularly by identifying when pupils reach particular levels of comprehension. In other words, although results are expressed in terms of fluency, fluency was secondary to comprehension in determining the benchmarks.

Table 13. EGRA Benchmarks for Reading (Passage) Fluency for Class 2[25]

	Emergent Reader	Fluent Reader
English	30 cwpm	65 cwpm
Kiswahili	17 cwpm	45 cwpm

Subsequently, in the process of developing a national meriting tool, the Ministry of Education, Science, and Technology set benchmark proficiency scores for mathematics in the early grades. These benchmarks are particularly useful for examining trends in absolute pupil performance.

Table 14. EGMA Benchmarks for Mathematics Fluency for Class 2[26]

	Emergent	Fluent
Addition	8 capm	12 cspm
Subtraction	8 capm	12 cspm

Using these benchmarks, we can see the change in the proportion of pupils at Bridge and other schools that have achieved fluency in literacy and numeracy at each grade level over time. Table 15 examines the additional percentage of pupils who met the "emergent" EGRA/EGMA benchmarks.

24 To our knowledge, the panel did not set benchmarks for other grade levels.
25 PRIMR Midline Report, p. 31. cwpm = correct words per minute.
26 PRIMR Final Report, p. 49. capm = correct addition items per minute, cspm = correct subtraction items per minute.

Table 15. Percentage Increase of Pupils Passing "Emergent" Benchmarks

(For All Pupils Tested in Both 2013 and 2014)

Early Grade Reading Assessment – English

	Overall		Pre-Unit to Class 1		Class 1 to Class 2		Class 2 to Class 3	
School Type	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Bridge	1287	30.97%	393	42.99%	423	36.03%	471	16.44%
Public	1430	23.83%	445	21.77%	484	34.15%	501	16.27%

Early Grade Math Assessment – Addition

	Overall		Pre-Unit to Class 1		Class 1 to Class 2		Class 2 to Class 3	
School Type	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Bridge	1286	24.04%	391	41.39%	426	23.59%	469	10.05%
Public	1449	19.86%	438	24.32%	505	26.74%	506	9.99%

Early Grade Math Assessment – Subtraction

	Overall		Pre-Unit to Class 1		Class 1 to Class 2		Class 2 to Class 3	
School Type	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Bridge	1287	9.66%	393	40.20%	423	23.22%	471	16.71%
Public	1430	2.77%	445	15.06%	484	17.02%	501	16.18%

Overall, in one year, Bridge moved a greater proportion of its pupils into meeting the "emergent" benchmarks for English reading, addition, and subtraction than public schools. This information is best used as an additional metric of Bridge's performance as a school system, as it is sensitive to the original percentage of pupils who were close to meeting the benchmark in the first place.[27]

Table 16 looks at the additional percentage of pupils who passed the fluency benchmarks, which is defined as proficiency at a Class 2 level. **Again, overall, a greater proportion of Bridge pupils have moved into the "fluent" category in one year.**

27 For example, take the case of School A that had X% of pupils that were right below the benchmark at baseline versus School B that had the same X% of pupils below the benchmark at baseline, but just further below the benchmark on an absolute level. The two schools may have created the same amount of gains for their pupils on average, but if School B's pupils still did not meet this absolute standard, School A's performance would look much better.

Table 16. Percentage Increase of Pupils Passing "Fluent" Benchmarks

(For All Pupils Tested in Both 2013 and 2014)

Early Grade Reading Assessment - English

	Overall		Pre-Unit to Class 1		Class 1 to Class 2		Class 2 to Class 3	
School Type	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
Bridge	1287	29.82%	393	6.11%	423	40.00%	471	40.39%
Public	1430	15.79%	445	2.25%	484	17.23%	501	26.40%

Early Grade Math Assessment - Addition

	Overall		Pre-Unit to Class 1		Class 1 to Class 2		Class 2 to Class 3	
School Type	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
Bridge	1287	20.65%	393	13.22%	423	26.40%	471	21.67%
Public	1430	17.61%	445	7.14%	484	21.19%	501	23.59%

Early Grade Math Assessment - Subtraction

	Overall		Pre-Unit to Class 1		Class 1 to Class 2		Class 2 to Class 3	
School Type	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
Bridge	1287	6.95%	393	5.85%	423	11.38%	471	12.67%
Public	1430	4.40%	445	3.37%	484	3.09%	501	12.67%

2. Cross-Country Comparison

In this section, we examine how Bridge's results in Kenya compare internationally to basic literacy and numeracy levels in other countries. For EGRA, differences in language structure and complexity preclude us from directly comparing average scores across countries. However, what *is* comparable is the proportion of pupils who scored a zero for each subtask – in other words, the proportion of pupils who could not read a single letter or word.[28] We reviewed EGRA reports for each country to find the appropriate statistics. Results for Class 2 by country and subtask are shown in Table 17 below.[29]

28 "The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy, Gove and Wetterberg." 2011, pg. 29.

29 Due to the time requirements of reviewing reports for every country participating in EGRA, we limited our research to countries in Africa.

Table 17. Class 2 Zero Scores for EGRA Subtasks by African Country

Country	Sample	Letter Sound	Listening Comp.	Non-word	Familiar Word	Passage Fluency	Reading Comp.
Kenya[30]	Bridge	--	21%	8%	3%	4%	14%
Kenya	Public	--	46%	13%	8%	11%	35%
Kenya (PRIMR)[31]	Treatment	--	--	--	--	7%	--
Kenya (PRIMR)	Control	--	--	--	--	13%	--
DRC[32]	Treatment	23%	55%	--	--	--	--
DRC	Control	19%	60%	--	--	--	--
Ghana[33]	All	51%	68%	75%	--	45%	85%
Tanzania[34]	All	38%	--	--	59%	38%	95%
Zambia[35]	All	50%	--	88%	--	91%	95%
Mozambique[36]	All	85%	--	--	83%	86%	86%
Malawi[37]	All	61%	--	96%	90%	96%	98%
Ethiopia[38]	All	--	--	--	--	34%	48%
Morocco[39]	All	18%	18%	34%		33%	62%

Across countries, there is wide variation in the proportion of pupils who cannot read a single word. While only 33% of Class 2 pupils in Morocco and 34% of Class 2 pupils in Ethiopia cannot read a single word, the same is true for 86% of pupils in Mozambique, 91% of pupils in Zambia, and 96% of pupils in Malawi at the same class level. In Kenya, the percentage of pupils unable to read a single word is relatively low, at 13% on average in non-PRIMR schools and only 7% in schools in the PRIMR program. By contrast, only 3% of Bridge International Academies' Class 2 pupils are unable to read a single word. **A lower proportion of pupils at Bridge International Academies cannot read a single letter, word, or passage when compared to their domestic and international peers.**

30 Note that for the pupil samples collected by the EGRA/EGMA project (shown in blue above), no scores are given for Letter Sound Knowledge because this subtask was not administered to pupils in Class 2.

31 Piper & Mugenda, 2014. "The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation." pg. 36

32 Bulat et al., 2012. "PAQUED: DRC. Projet d'Amelioration de la Qualite de l'Education (PAQUED). Midterm Report, Early Grade Reading Assessment (EGRA)." pg. 16

33 Results were obtained by averaging scores for urban and rural pupils for English EGRA results. Kochetkova & Brombacher, 2014. "Ghana 2013 Early Grade Reading Assessment and Early Grade Mathematics Assessment: Report of Findings." pg. 16, 15, 17, 18, 19.

34 Brombacher et al., 2014. "National Baseline Assessment for the 3Rs (Reading, Writing, and Arithmetic) using EGRA, EGMA, and SSME in Tanzania: Study Report." pg. 26

35 Collins et al., 2012. "Pupil Performance, Pedagogic Practice, and School Management: An SSME Pilot in Zambia." pg. 38, 39, 40.

36 Raupp et al., 2013. "Impact Evaluation for the USAID / Aprender A Ler Project in Mozambique. Baseline Report." Scores were calculated by averaging the percentages for Nampula and Zambézia. pg. 29, 31, 33, 34

37 Miksic, Emily and Stephen Harvey, 2012. "Malawi National Early Grade Reading Midterm Assessment 2011." pg. 18

38 Piper, Benjamin, 2010. "Ethiopia Early Grade Reading Assessment. Data Analytic Report: Language and Early Learning." pg. 21 and 22

39 Messaoud-Galusi et al., 2012. "Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management in Doukkala Abda, Morocco." pg. 34

For EGMA, two independent sources at RTI (the organization that developed the assessment) strongly discouraged international comparison of subtasks. In particular, RTI's math experts note that "EGMA was not designed for international comparisons." They also underscore recent findings that indicate that "language of instruction, home language, and home environment appear to affect outcomes much more than we would think for math." However, if we must perform international comparisons, the best way would be to examine zero scores.

Therefore, we present zero scores for EGMA across countries. Table 18 below compares zero scores in Class 2 for Bridge with those presented in studies conducted in Kenya and in other African countries. Note that while 8% of Class 2 pupils in Ghana cannot successfully complete a single basic addition problem, the same is true for 22% of Class 2 pupils in Morocco. 15% of Bridge Class 2 pupils are unable to do so. For Subtraction, the percentage of Class 2 pupils who cannot answer a single problem correctly in one minute is even higher, with 19% of Ghanaians and 44% of Moroccans unable to complete a single subtraction problem. For Bridge, the figure is 11%. **Again, a lower proportion of pupils at Bridge International Academies cannot perform basic numeracy skills when compared to their international peers.**[40]

Table 18. Class 2 Zero Scores for EGMA Subtasks by Country

Country	Sample	Quantity Disc.	Add 1	Add 2	Sub 1	Sub 2	Word Problems
Kenya[41]	Bridge	--	15%	1%	11%	1%	18%
Kenya	Public	--	24%	2 %	12%	4%	27%
DRC[42]	Treatment	2%	19%		35%		35%
DRC	Control	3%	20%		36%		27%
Ghana[43]	All	4%	8%	51%	19%	70%	9%
Tanzania[44]	All	5%	12%	48%	22%	58%	24%
Zambia[45]	All	19%	21%	--	26%	--	--
Morocco[46]	All	5%	22%	--	44%	--	--

40 We were unable to locate zero scores for EGMA subtasks in the PRIMR endline or final reports.

41 Note that for the pupil samples collected by the EGRA/EGMA (shown in blue above), no scores are given for Quantity Discrimination because this subtask was not administered to pupils in Class 2.

42 Bulat et al., 2012. "PAQUED: DRC. Projet d'Amelioration de la Qualite de l'Education (PAQUED). Midterm Report, Early Grade Reading Assessment (EGRA)." pg. 16

43 Results were obtained by averaging scores for urban and rural pupils for English EGRA results. Kochetkova & Brombacher, 2014. "Ghana 2013 Early Grade Reading Assessment and Early Grade Mathematics Assessment: Report of Findings." pg. 21

44 Brombacher et al., 2014. "National Baseline Assessment for the 3Rs (Reading, Writing, and Arithmetic) using EGRA, EGMA, and SSME in Tanzania: Study Report." pg. 32

45 Collins et al., 2012. "Pupil Performance, Pedagogic Practice, and School Management: An SSME Pilot in Zambia." pg. 45 and 47

46 Messaoud-Galusi et al., 2012. "Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management in Doukkala Abda, Morocco." pg. 42

3. Effect Sizes of Other Educational Interventions

We examine effect sizes from other educational interventions in Table 19 below. Despite each study's operation in very different contexts, comparability is maintained with the common adoption of rigorous evaluation methods and the use of effect sizes for reporting results.

We find that other major educational interventions yield effect sizes ranging from 0.17 to 0.46. It is noteworthy that nearly all of these interventions, with the exception of the KIPP Middle Schools study,[47] are randomised controlled trials. Therefore, their effect sizes can be interpreted as causal impact. Although the KIPP Middle Schools study was not an RCT, the study design is rigorous enough that its effect sizes approach an impact estimate.

Table 19. Effect Sizes for Selected Education Interventions

Study (Country)	Study Design & Intervention	Time Between Pre/Post-Test	Effect Size – Reading	Effect Size – Math
KIPP Middle Schools (U.S.A.)[48]	Matched comparison study; enrolment in a KIPP charter school	3 years	0.21	0.36
Teacher Performance Pay (India)[49]	RCT; receipt of group bonus based on school performance or receipt of Individual bonus based on teacher performance	2 years	0.17	0.27
Merit Scholarship Program for Girls[50] (Kenya)	RCT; receipt of scholarship covering school fees and cash grant for school supplies	2 years	0.2 to 0.3 (across multiple subjects, including English and math)	
Primary Math and Reading (PRIMR) Initiative[51] (Kenya)	RCT; provision of low-cost books to each pupil, targeted lesson plans, training on lesson delivery, and instructional support	2 years	0.46 (on selected subtasks)	0.2 (across all subtasks)

The closest comparison to our study is the PRIMR study, though there are a few key differences. First, while they do follow a cohort of 600 pupils through time, these results are not reported publicly at this time. Therefore, the effects presented here are from their difference-in-differences study on the same schools throughout time.[52] Second, the study is across two years and looks at Class 1 and 2 classrooms; ours spans a single year and tracks pupils beginning in Pre-Unit, Class 1, and Class 2 and ending in Class 1, Class 2, and Class 3. Third, these figures include the effects of the programme at both public and low-cost private schools, where effects from the programme implemented in low-costs private schools were actually greater. Finally, the average effect size reported in the table above for PRIMR reflects a different set of subtasks than we use.

47 Tuttle et al. (2012)

48 Tuttle et al (2013)

49 Muralidharan and Sundararaman (2011)

50 Kremer et al. (2004)

51 Piper and Mugenda (2014)

52 They do collect this information, but perhaps do not use it in their methodology given the RCT design.

For the subtasks specifically reported, we compare their 2-year overall results with our 1-year overall results – again with the caveat that RTI has employed a different methodology[53] and the grade levels differ. In addition, we report all subtasks with statistically insignificant effect sizes as zeros instead of omitting them, which drastically reduces the magnitude of our overall average effect sizes.

Table 20. Apples-to-Oranges Comparison of Treatment Effects

	EGRA/EGMA	PRIMR
Subtask	**Panel DiD, with Controls**	**Repeated Cross Section, DiD**
Listening Comprehension	0.49	N/A
Non-word Reading	0.15	N/A
Familiar Word Reading	0.24	N/A
Passage Fluency	0.31	0.40
Reading Comprehension	0.39	0.38
Quantity Discrimination	0.32	0.03
Addition Level 1	0	0.17
Addition Level 2	0	N/A
Subtraction Level 1	0	0.21
Subtraction Level 2	0	N/A
Word Problems	0.20	0.13
EGRA Effect Size, p<.05	**0.31**	**0.39**
EGMA Effect Size, p<.05	**0.09**	**0.14**

When we limit PRIMR results to include only the subtasks we also administered, the 2-year EGRA PRIMR effect size for English is reduced to .39 and for maths is reduced to .14 on average.

53 While effect sizes do better allow us to compare impacts of various treatments across studies, there likely remain differences in effect sizes within a study due to the selection of the evaluation model, econometric specifications, geographic locations, and inclusion of different school types as a comparison group.

VII. Limitations and Robustness Checks

Despite these improvements that make our study one of the most credible among education or development organizations, we acknowledge that there are potential threats to its validity. Our main concern is attrition – 41% of pupils in our evaluation drop out before the study's completion, making it impossible to collect data on their outcomes. This is due to a variety of reasons, many stemming from DMC's ability to find and assess pupils at comparison schools. We are constrained by the schedules of comparison schools, the willingness of headmasters to participate, as well as whether or not pupils happen to be absent on a given day of testing. We are unable to track pupils outside of their baseline school because doing so is cost prohibitive.

Two types of attrition can occur in a study: equivalent attrition and differential attrition. Equivalent attrition occurs when individuals from the groups we are comparing attrite, but each group's composition remains the same. While this type of attrition does not lead to systematic differences between the treatment group and the control group, it often raises concerns at high levels. The inclusion criteria for behavioural intervention studies, for example, require retention rates of at least 70%. The CDC follows a similar logic, requiring retention to be at least 70% for an intervention to be classified as "effective" and 60% for "promising."

Differential attrition occurs when attrition patterns are different for a particular type of pupil or vary across comparison groups. The former is a concern because it results in inaccurate measures of the magnitude of an intervention's effect, despite there being no systematic differences between comparison groups. The latter is a concern because it causes systematic differences in the characteristics between comparison groups, which ultimately can lead to biased estimates of results.[54] Furthermore, if specific types of individuals are more likely to attrite, it means that the sample now looks different from the population it was meant to represent, and the study's results may no longer be generalizable to the larger context.[55] A study in the American Journal of Public Health points out that the real concern is not necessarily high attrition itself, which frequently happens in highly mobile populations, but differential attrition.[56] The question therefore becomes whether there is *differential* attrition and, if there is, what we can do about it to make our findings more robust.

54 What Works Clearinghouse, v2.1
55 Miller and Hollist (2007)
56 Amico (2009)

1. Differential Attrition

Differential attrition has critical implications for how analyses and results are examined. Both types of differential attrition are potential threats to the validity of our study:

1. *Is a particular type of pupil (e.g., high- vs. low-performing) leaving our sample overall?* Say, for instance, that low-performing pupils are the ones dropping out of the sample, such that the effects we find are driven by more high-performers remaining in the sample. Does this mean that any effects we find would not be generalizable to all types of pupils?

2. *Are different types of pupils (e.g., high- vs. low-performing) leaving our sample at different rates depending on school type?* For example, if a significantly greater proportion of high-performing pupils leave public schools and these same pupils are more likely to make greater gains over time, the estimate of Bridge's effect would be biased in our favour. This is because we, unlike public schools, are left with a higher-performing sample of pupils over time.

To get a sense of whether or not we should be concerned about each type of differential attrition, we return to summary statistics to examine whether differences between pupils who stayed and left the sample are similar across school types. The tables are detailed in Appendix 7.

For Bridge, a greater proportion of attriting pupils are older and male. More pupils who left the sample had zero scores, and for those in the oldest cohort who began in Class 2, had lower beginning levels of reading fluency and comprehension. For EGMA, however, the story on zero scores is mixed depending on the skill assessed. For higher level maths, those who attrited tended to have lower scores.

For the public schools, a greater proportion of attriting pupils are again older and did not attend ECE, but there are no statistically significant gender differences. 1% more pupils who stayed in the sample had breakfast. In terms of household characteristics, a greater proportion of non-attriting pupils' households have a radio, television, and electricity. These pupils were also more likely to enjoy the radio and television with their parents. Again, pupils who stayed in the sample did better on subtasks at baseline than those who did not.

2. Tests for Differential Attrition

The previous section offered a simple exploration of differential attrition: Across school types, older and lower performing pupils attrite more heavily. While it is less clear whether different types of pupils are attriting from different types of schools between administration rounds, differences in attrition levels based on baseline test scores warrant further investigation. Therefore, in the following sections, we formally test for differential attrition.

a. Attrition Probits

The most straightforward way to conduct a test for non-random attrition is to examine the probability of attrition given information we have on both attrited and non-attrited pupils.[57] We ran a series of probit regressions that consider the predictive power of pupil baseline information on the likelihood of attrition. Each regression also tests to see if such baseline information predicts attrition differently at Bridge than at other schools. Specifically, for each subtask, we ran the following model:

$$attrite = \beta_0 + \beta_1(Bridge) + \beta_2(baseline\ subtask\ score) + \beta_3(baseline\ subtask\ score*Bridge) + \beta_A(baseline\ characteristics) + \beta_B(baseline\ characteristics*Bridge) + \beta_C(county\ dummies) + \beta_D(county\ dummies*Bridge) + \varepsilon$$

This comprehensive probit model considers a number of factors, including: (1) school type; (2) pupil baseline test score; (3) the interaction of pupil baseline test score and school type; (4) a set of baseline characteristics, which includes: age, gender, years of ECE; whether the pupil has a radio, television, or electricity in the household; whether the pupil speaks their mother tongue or English at home; whether the pupil listens to the radio or watches television; whether the pupil is read to or receives homework help at home; the meal count of the pupil within a day of testing, along with the pupil's diet diversity score (DDS); (5) the interaction of this set of baseline characteristics with school type; (6) county location; and (7) county location interacted with Bridge. Note that factors (1), (2) and (4) test the first type of attrition, which considers whether different types of pupils (i.e., high-performing vs. low-performing) are more likely to attrite from our sample. (3) and (5) test the second type of attrition, which considers whether different types of pupils (i.e., high-performing vs. low-performing) from a particular school type are more likely to attrite from our sample than another school type.

The results of these probit regressions are in Appendix 8.1, Table 36.

First, we examine differences in the likelihood of attrition based on their characteristics in 2013 across all pupils in our sample.

We find that pupils attending Bridge are more likely to attrite from the sample. Because we did not assess pupils if they were not allowed in class, the high attrition rates could be partially due to parents' inability to pay. However, the primary reason for attrition at non-Bridge schools is probably the lack of permission to assess pupils or lack of time to assess all pupils. **Pupils with higher baseline test scores are generally less likely to attrite, as are girls.** Pupils who speak their mother tongue and English at home are more likely to attrite, but since the vast majority of pupils speak their mother tongue at home, it is unclear what the take-away should be from this.

Second, we dive into the key question of whether or not there is differential attrition by school type *and* baseline characteristics – in other words, are pupils of a certain "type" more or less likely to attrite from Bridge than from other schools?

57 According to the Chronic Poverty Research Centre (CPRC) supported by DFID, "the simplest test for whether attrition is random is to estimate a probit in which the dependent variable takes the value one for households which drop out of the sample after the first wave (attrite) and zero otherwise. Explanatory variables are baseline values for all variables that are believed to affect the outcome variable of interest plus any available variables which characterise the interview process. It is usual to include lagged values of the outcome variable in such attrition probits. As pointed out by Outes-Leon and Dercon (2008), it is also useful to examine the pseudo R-squared from attrition probits, as they can be interpreted as the proportion of attrition that is non-random." (Baulch, B. and Quisumbing, A., 2011)

In comparison to public schools, we find that there is no difference in attrition rates at Bridge based on pupils' baseline performance. Troublingly however, while girls in general are less likely to attrite from the sample, girls at Bridge are more likely to attrite. If a Bridge pupil had household assets or spoke their mother tongue or English at home, they were less likely to attrite. Attrition by location is also significant.

Third, we also examine whether any of the learning activities at home (receiving homework help or being read to) were jointly significant. It's possible that individual indicators don't predict attrition but that, jointly, the family of indicators is significant. However, in this case, the group of learning characteristics still were not significant. Neither were indicators of nutrition based on the number of meals and diet diversity score in a joint test.

In short, there is differential attrition on observables when we compare Bridge to public schools, though not on the key variables of interest – baseline test scores – which we believe is the best predictor of across pupil differences.

b. Pooling Attrition

We also conduct a Becketti, Gould, Lillard & Welch (BGLW) test to see if baseline characteristics and test scores differ for those who subsequently attrite or stay in the sample, and additionally, whether they are different both across attrition status and school type.[58] We create (1) interactions between the attrition indicator and each baseline variable along with (2) additional interaction terms with Bridge to run the following specification:

$$\textit{baseline subtask score} = \beta_0 + \beta_1(\textit{Bridge}) + \beta_2(\textit{attrite}) + \beta_A(\textit{baseline characteristics}) + \beta_B(\textit{county dummies}) + \beta_C(\textit{baseline characteristics*Bridge}) + \beta_D(\textit{county dummies*Bridge}) + \beta_4(\textit{attrite*Bridge}) + \beta_E(\textit{attrite*baseline characteristics}) + \beta_F(\textit{attrite*county dummies}) + \beta_G(\textit{attrite*baseline characteristics*Bridge}) + \beta_H(\textit{attrite*county dummies*Bridge}) + \varepsilon$$

We then conduct an F-test to see if the attrition dummy with the interaction terms are all jointly equal to zero. The idea is that while any one interaction term may be insignificant, it could be significant in conjunction with others. When looking at Bridge and public schools, the F-test is significant for 4 out of 5 EGRA subtasks and 2 out of 6 EGMA subtasks, which means there likely is differential attrition – i.e., there is a difference in baseline performance of different types of pupils who attend different types of schools. While we do control for baseline test scores, this finding gives us pause and suggests that additional robustness checks are warranted. More details can be found in Appendix 8.2.

3. Robustness Checks

The best way to deal with differential attrition is to successfully track down and assess pupils who attrited from the sample. The goal would be to assess the "attritors", and conduct a parallel study to examine whether the impact of Bridge still holds. Unfortunately, however, it would require funds at least ten times our current budget in order to do the exercise well. Furthermore, because we did not collect detailed contact information during the time of the assessment, successfully following pupils who have moved neighborhoods would be an extremely challenging endeavor.

While we are unable to tackle the high attrition rates head on, we conducted two types of robustness checks: a fixed-effects analysis and a bounds analysis using Lee bounds. Fixed effects control for observable and unobservable characteristics of the pupils in the sample. Lee bounds gives us a range of effects based on varying assumptions about attrition.

a. Panel Fixed-Effects Model

We conducted an additional analysis using panel data with fixed effects to address the fact that we can only control for observed values in the panel DiD model. Panel with fixed effects (FE) is a powerful model because unlike the simple panel differences-in-differences with controls, it controls for unobservable characteristics. Many indicators, such as a pupil's socioeconomic status or innate ability, are incredible difficult to obtain in an accurate way. The fixed effects model captures all of this variation between individual pupils and finds a more precise effect.

However, we cannot rely on the Panel FE entirely because it allows for less flexibility than the panel with controls. In panel with controls, we controlled for baseline test score as well as its squared and cubed terms, allowing for nonlinear growth. Fixed effects forces growth into a linear model, meaning that it assumes the lowest-scoring pupil, the median pupil, and the highest-scoring pupil all are assumed to have the same growth trajectory. Therefore, we will interpret the panel with controls as an upper bound for the Bridge effect and the panel FE as a lower bound.

Compared to public school pupils, Bridge pupils experience anywhere from a .19 to a .31 standard deviation gain in English scores on average in one academic year. In maths, the lower bound of the effect size is .04 standard deviations lower, with an upper bound of .09 standard deviations higher. For more details about fixed effects, see Appendix 5.

b. Lee Bounds

Lee bounds computes a lower and upper bound of the Bridge treatment effect by trimming the dataset at the extremes based on what the data says. Specifically, its exclusion criteria is based on a few factors: the outcome of interest, attrition rates, and school type. For example, assume we believe that relatively more high performing pupils from other schools leave and bias the sample in our favour. Applying Lee bounds would trim our sample so that we get a lower bound estimate of what our effects would be.

Unfortunately, because Lee bounds makes assumptions relying on attrition rates themselves among other factors, the high absolute level of attrition in our sample rendered this robustness check unhelpful. The range established by our Lee bounds estimate was too wide for it to add any meaningful insights on the validity of our estimates.

58 If we had this information, there is a second test for non-random attrition we could have conducted. The CPRC points to the "The Becketti, Gould, Lillard and Welch (hereafter BGLW) test [,which] involves regressing an outcome variable from the first wave of a survey on household and community variables, an attrition dummy, and the attrition dummy interacted with the other explanatory variables. An F-test of the joint significance of the attrition dummy and the interaction variables is then conducted to determine whether the coefficients from the explanatory variables differ between households who are retain or attrite from the panel." See Testing and adjusting for attrition in household panel data, CPRC Toolkit Note.

VIII. Conclusion

As an organization seeking to constantly improve learning outcomes for its pupils, Bridge relies on rigorous evidence and continuous feedback loops to improve programmatic efforts. This study represents a significant improvement from previous internal studies on several fronts. First, as Bridge expanded, we increased the number of schools in our study to ensure an adequate sample size. We also stratified our sampling by former province in order to get a nationally representative sample. Next, we bolstered our survey instrument and tightened our data collection and monitoring strategies. Finally, we applied more rigorous methods and robustness checks to our data analysis.

Ultimately, we want to learn whether or not what we do works. Randomised evaluations are the gold standard for providing such evidence, but like other organizations, Bridge faces external constraints that prevent us from implementing an RCT to study our own impact relative to neighbouring schools. As such, we have agreed to participate in a randomised evaluation of Bridge led by a team of external researchers with significant experience in Kenya and its education system. We anticipate the study to begin in 2016.

In the meantime, Bridge has pursued the next best thing, which is this quasi-experimental study of early grade primary pupils in over 87 schools across Kenya. We look forward to your feedback.

References

Baroody, A. J., & Wilkins, J. L. M. (1999). “The development of informal counting, number, and arithmetic skills and concepts.” In J. V. Copely (Ed.), Mathematics in the early years (48–65). Reston, Virginia: National Council of Teachers of Mathematics.

Battista, J. (1990). Spatial visualization and gender differences in high school geometry. Journal of Research in Mathematics Education, 21; 47-60.

Baulch, B., & Quisumbing, A. (2011). Testing and adjusting for attrition in household panel data. Toolkit Note. Available at http://www.chronicpoverty.org/publications/ details/testing-and-adjusting-for-attrition-in-household-panel-data

Bloom, H.S. et al. (2008). Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions. Journal of Research on Educational Effectiveness, 1; 289-328.

Bloom, H.S. et al (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. Child Development Perspectives, 2(3); 172-177

Brombacher, A., Nordstrum, L., Davidson, M., Batchelder, K., Cummiskey, C., King, S., & The Research Triangle Institute (2014). USAID/Tanzania: National Baseline Assessment for the 3Rs (Reading, Writing, and Arithmetic) Using EGRA, EGMA, and SSME in Tanzania. Accessed May 7, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=TZ&statusID=3&showtypes=0

Buckley, J. & Shang, Y. (2003). Estimating policy and program effects with observational data: the ‘differences-in-differences’ estimator. Practical Assessment, Research & Evaluation, 8(24). Available at http://PAREonline.net/getvn.asp?v=8&n=24

Bulat, J., Brombacher, A., Slade, T., Iriondo-Perez, J., Kelly, M., Edwards, S., & The Research Triangle Institute (2014). USAID/DRC: PAQUED 2014 Endline Report of Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA). Accessed May 7, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=CD&statusID=3&showtypes=0

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. Journal for Research in Mathematics Education, 15(3); 179–202.

Carpenter, T. P., Hiebert, J., & Moser, J. M. (1981). Problem structure and first-grade children’s initial solution processes for simple addition and subtraction problems. Journal for Research in Mathematics Education, 12(1); 27–39.

Casey, B., (2004). “Mathematics problem-solving adventures: A language-arts-based supplementary series for early childhood that focuses on spatial sense.” In D. H. Clements & J. Sarama (Eds.), Engaging young children in mathematics: Standards for early childhood mathematics education (377–392). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. School Psychology Review, 33(2); 234–248.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in pupil outcomes. Remedial and Special Education, 29(1); 46–57.

Clements, D. H. (2004). Geometric and spatial thinking in early childhood education. In D. H. Clements & J. Sarama (Eds.), Engaging young children in mathematics: Standards for early childhood mathematics education (pp. 267–297). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. et al. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review In Review of Educational Research 66(3): 227-268.

Cooper, R. G., Starkey, P., Blevins, B., Goth, P., & Leitner, E. (1978). Number development: Addition and subtraction. Paper presented at the 8th annual meeting of the Jean Piaget Society, Philadelphia, May 1978.

Cremata, E. et al (2013). National Charter School Study. Stanford: Center for Research on Education Outcomes.

Floyd, R. G., Hojnoski, R., & Key, J. (2006). Preliminary evidence of the technical adequacy of the preschool numeracy indicators. School Psychology Review, 35(4), 627–644.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. School Psychology Review, 33(2); 188–192.

Gelman, R., & Gallistel, C. R. (1986). “The child’s understanding of number.” Cambridge, Massachusetts: Harvard University Press.

Gong, E. (2009). Differences-in-differences [PowerPoint presentation]. Available at www.ocf.berkeley.edu/~garret/decal/dnd_slides09.pptx.

Grande, J. D., & Morrow, L. (1993). “Curriculum and evaluation standards for school mathematics: Grades K–6.” Reston, VA: The National Council of Teacher of Mathematics.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. The Reading Teacher, 59(7); 636–644.

Headey, D., & Ecker, O. (2013). Rethinking the measurement of food security: from first principles to best practice. Food Security, 5(3); 327-343. doi: 10.1007/s12571-013-0253-0

Jones, S., & Uwezo (2012). “Are our Children Learning? Literacy and Numeracy Across East Africa, August 2012.” Retrieved from http://www.uwezo.net/wp-content/uploads/2012/09/RO_2012_UwezoEastAfricaReport.pdf

Jones, S., & Uwezo (2013). “Are our Children Learning? Annual Assessment Report.” Retrieved from http://www.uwezo.net/wp-content/uploads/2012/08/2013-Annual-Report-Final-Web-version.pdf

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. Child Development, 74(3); 834–850.

Kersh, J., Casey, B. M., & Young, J. M. (2007). “Research on spatial skills and block building in girls and boys: The relationship to later mathematics learning.” In O. N. Saracho & B. Spodek (Eds.), Contemporary perspectives on mathematics in early childhood education (233–251). Charlotte, North Carolina: Information Age Publishing.

Klein, A., & Starkey, P. (2004). “Fostering preschool children’s mathematical knowledge: Findings from the Berkeley Readiness Project.” In D. H. Clements & J. Sarama (Eds.), Engaging young children in mathematics: Standards for early childhood mathematics education (343–360). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kochetkova, E., Brombacher, A., & Research Triangle Institute (2014). USAID/Ghana: Early Grade Reading Assessment and Early Grade Mathematics Assessment Report of Findings. Accessed May 7, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=GH&statusID=3&showtypes=0

Kremer, M. et al. (2004). Incentives to Learn. Available at http://www.nber.org/papers/w10971

McCombs, J.S. et al. (2011). Making Summer Count: How Summer Programs Can Boost Children’s Learning. Available at http://www.rand.org/content/dam/rand/pubs/monographs/2011/RAND_MG1120.pdf

Miller, R., & Hollist, C. (2007). Attrition bias. Faculty Publications, Department of Child, Youth, and Family Studies. Lincoln: Digital Commons at the University of Nebraska – Lincoln. Available at http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article =1044&context=famconfacpub

Muralidharan, K. & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. Journal of Political Economy 119(1); 39-77.

NCTM. (2008). 2007 TIMSS shows continued improvement in math. Retrieved January 1, 2009, from http://www.nctm.org/news/ content.aspx?id=17021

Piper, B., Mugenda, A., & Research Triangle Institute (2013). USAID/Kenya: The Primary Math and Reading (PRIMR) Initiative: Midline Impact Evaluation. Accessed May 7, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=KE&statusID=3&showtypes=0

Piper, B., Mugenda, A., & Research Triangle Institute (2014). USAID/Kenya: The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation. Accessed May 11, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=KE&statusID=3&showtypes=0

Piper, B., Mugenda, A., & Research Triangle Institute (2014). USAID/Kenya: The Primary Math and Reading (PRIMR) Initiative: Final Report. Accessed May 11, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=KE&statusID=3&s howtypes=0

Pouezevara, S., Sock, M., & Ndiaye, A. (2010). Evaluation des compétences fondamentales en lecture au Sénégal [Early grade reading assessment for Senegal]. Report prepared for the William and Flora Hewlett Foundation. Research Triangle Park, North Carolina: RTI International and Focus Africa. Retrieved September 17, 2010, from USAID website: https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=250

The Research Triangle Institute (2014). USAID/Nigeria: Results of the 2014 Hausa and English Early Grade Reading Assessments (EGRAs) in Government Primary Schools and IQTE Centers of Jigawa, Kaduna, Kano, and Katsina States. Accessed May 7, 2015. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=NG&statusID=3&showtypes=0

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. Cognition and Instruction, 5(1), 49–101.

Schneider, A. et al. (2010). Linear Regression Analysis Dtsch Arztebl Int. 107(44): 776–782. Mainz: Johannes Gutenberg University. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2992018/

Tuttle, C.C. et al (2013). KIPP Middle Schools: Impacts on Achievement and Other Outcomes. Washington, D.C.: Mathematica Policy Research.

Valentine, J. C., & Cooper, H. (2003). Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes. Washington, DC: What Works Clearinghouse.

What Works Clearinghouse (2013). Assessing Attrition Bias. Institution of Education Sciences. Washington, DC: What Works Clearinghouse.

What Works Clearinghouse (2008). Procedures and Standards Handbook (Version 2.0). Washington, DC: What Works Clearinghouse.

Appendix

A1. Randomization Strategy

To ensure that our samples are randomised in a way that takes into account gender proportions for given classes, team leaders will follow the six steps detailed below.

The most important numbers to keep in mind for this sampling strategy are the **interval** by which pupils will be selected and the **percentages of girls and boys**.

- **Separate Pupils by Gender:** Ask pupils to form two lines – one line for boys and one line for girls. If there are multiple streams for a class, collect *all* pupils and separate them out by boys and girls.

- **Determine the Total Number of Pupils for the Class:** Count the number of girls and the number of boys to find the total number of pupils. Example:

  Number of Girls = 5
  Number of Boys = 10
  Number of Girls + Number of Boys = 5 + 10 = 15
  We have a total of 15 pupils in the class we are assessing.

- **Calculate the Target Interval Number:** Take the **total number of pupils** and divide it by the target number of pupils. Example:

  Total number of pupils = 15
  Target sample number = 5

$$\frac{\text{Total Number of Pupils}}{\text{Target Sample Number}} \quad \frac{15}{5} = 3$$

  Based on the equation above, we find that our **target interval number** is 3.

- As a reminder, you want to oversample so you have a few alternates. We recommend sampling 22 per class for Pre-Unit, Class 1, and Class 2. We recommend sampling 12 for nursery. This is particularly important for Bridge pupils, as some of them may have “Not Allowed In Class” status due to prolonged non-payment of fees and therefore they cannot be assessed.

- **Calculate the Gender Ratios:** Calculate the ratio of girls to the total number of pupils and the ratio of boys to the total number of pupils in the class being assessed. To determine the ratio of girls, divide the number of girls in the class by the total number of pupils.

$$\frac{\text{Number of Girls}}{\text{Total Number of Pupils}} \quad \frac{5}{15} = \frac{1}{3} \blacktriangleright \quad \text{Ratio of Girls} = 1/3$$

$$\frac{\text{Number of Boys}}{\text{Total Number of Pupils}} \quad \frac{10}{15} = \frac{2}{3} \blacktriangleright \quad \text{Ratio of Boys} = 2/3$$

- **Calculate Number of Boys and Girls Required for the Sample:** Multiply the ratios found in step 4 by the target sample number to determine exactly how many girls and how many boys should be in the final sample. If the calculation yields a decimal, keep the following in mind: if the decimal is less than 0.5, round down. If the decimal is 0.5 or higher, round up. Example:

  Girls: Multiply the ratio of girls ($^{1}/_{3}$) by the target sample number (5).
  So: $^{1}/_{3}$ * 5 = 1.7. We should have 2 girls in our final sample.

  Boys: Multiply the ratio of boys ($^{2}/_{3}$), by the target sample number (5).
  So: $^{2}/_{3}$ * 5 = 3.3. We should have 3 boys in our sample.

- **Identify the Pupils for the Sample:** Use the interval to separately identify the girls and boys who will be in the sample and be assessed. See the tables below for which pupils would be selected, based on our calculations in steps 1 through 5. Our interval number was 3, so count every third girl and third boy. For the girls, once you've reached the fifth girl (who would be 2), start back at the top.

Random Selection, Example 1

Girls	
G	**2nd Selected Girl**
G	
G	**1st Selected Girl**
G	
G	

Boys	
B	
B	
B	**1st Selected Boy**
B	
B	
B	**2nd Selected Boy**
B	
B	
B	**3rd Selected Boy**
B	

Other Considerations:

What if the class being assessed has fewer than 18 pupils?

Follow the steps below to ensure that the *order* in which pupils are assessed is still random:

- Separate out the pupils by gender into two lines.
- Count every 3rd pupil and write that pupil's name down, until you have written down all pupil names for both lines. This line up represents the final order that pupils will be assessed.
- Example: Let's say there are 10 pupils: 5 boys and 5 girls. Even though we know we will have to assess all of them, if we just wrote their names down in the order they stood in line and assessed them, it would not be random (maybe sleepier kids are all at the end of the line). See the table below for the order:

Random Selection, Example 2

Girls	
G	2nd Selected Girl
G	4th Selected Girl
G	1st Selected Girl
G	5th Selected Girl
G	3rd Selected Girl

Boys	
B	2nd Selected Boy
B	4th Selected Boy
B	1st Selected Boy
B	5th Selected Boy
B	3rd Selected Boy

A2. EGRA & EGMA Subtask Descriptions

1. Listening Comprehension

On the Listening Comprehension task, the assessor reads a passage aloud that narrates a familiar activity or event, and then asks the pupil to answer questions about what the pupil has just heard. The task tests listening comprehension, a skill that is separate from reading comprehension, as learners approach, process, and respond to speech and text in different ways. As in oral vocabulary, poor performance on a listening comprehension task suggests that the child may not have the vocabulary that the reading materials use, the child has difficulty processing what is heard, or both. Both direct factual questions (answers that are directly available in the passage) and inference questions (answers that require additional information, insight, or knowledge not directly available from the text) can be asked. This task was administered to pupils in Nursery, Pre-Unit, Class 1, and Class 2.

2. Non-word Reading

The Non-word Reading task (also sometimes called the "Invented" or "Unfamiliar Words" task) measures a pupil's decoding ability and is designed to detect problems of sight recognition of words. Examples in English include "fet," "ca," "ut," and "bleb." Many children in the early grades learn to memorize or recognize a broad range of "sight words" (words that primary school children are taught to recognize on sight, as many of these words are not easy to sound out phonetically, and thus must be memorized). Successful readers avoid memorization of text and combine both decoding and sight-recognition skills. Testing for how well a child can decode invented words provides a better estimate of the child's ability to read unfamiliar words that fall outside his/her sight recognition vocabulary. Pupils are asked to read as many "non-words" as they can in one minute; the assessor then records the number of non-words the pupil read correctly. If the pupil reads all of the non-words before 60 seconds have passed, the assessor notes the amount of time remaining. This task was administered to pupils in Nursery, Pre-Unit, Class 1, and Class 2.

3. Familiar Word Reading

The Familiar Word Reading task tests pupils' ability to read simple, common one- and two-syllable words. One way to examine reading fluency is by assessing how well a pupil can read a paragraph. However, for the purposes of measuring word recognition and decoding skills, a better method is to assess how well the pupil can read a list of unrelated words. That way, the pupil cannot simply guess the next word

from the context provided by surrounding words. For this assessment, familiar words are common words selected from early grade reading materials and storybooks for first-, second-, and third-grade materials (progressively increasing in difficulty). Pupils are asked to read as many familiar words as they can in one minute; the assessor times the pupil and records the number of correct familiar words read per minute. If the pupil reads all of the non-words before 60 seconds have passed, the assessor notes the amount of time remaining. This task was administered to pupils in Nursery, Pre-Unit and Classes 1 to 4.

4. Passage Fluency

The Passage Fluency task requires that pupils read a text with high accuracy and good speed. This task measures overall reading competence: the ability to translate letters into sounds, unify sounds into words, process connections, relate text to meaning, and make inferences to fill in missing information.[59] As skilled readers translate text into spoken language, they combine these tasks in a seemingly effortless manner; because oral reading fluency captures this complex process, it can be used to characterize reading skill.[60] The pupil is given one minute to read the passage. This test was administered to pupils in grades Pre-Unit through Class 4.

5. Reading Comprehension

The Reading Comprehension task requires pupils to respond correctly to different types of questions, including literal and inferential questions about the text they have read aloud. Assessors ask pupils to read a passage, stopping them after 1 minute and recording the number of words read. Then, pupils are asked comprehension questions that include direct, fact-based questions, as well as at least one question requiring inference from the text. Poor performance on a reading comprehension task would suggest that the pupil has trouble with decoding, or with reading fluently enough to comprehend, or with vocabulary. This task was administered to pupils in grades Pre-Unit through Class 4.

Please note that this analysis calculates the percentage of correct answers out of a total of five questions. However, the number of correct answers a pupil can provide depends on how far in the passage he or she read in the previous task, Passage Fluency. If a pupil did not read quickly enough to finish the entire passage in the previous task, then only some of the 5 comprehension questions were not administered: pupils were asked as many questions in Reading Comprehension as were answerable, given the number of sentences read in the Passage Fluency task. Pupils were therefore only asked the questions that corresponded with the portion of the passage that they were able to read during the timed exercise. While this scoring method differs from how reading comprehension is generally scored, upon consultation with our academic team, we believe this method more clearly represents differences in pupil learning.

6. Quantity Discrimination

The Quantity Discrimination task measures children's ability to make judgments about differences by comparing quantities. This skill is tested using numbers or objects, such as circles, and asking which group has more. Quantity discrimination in Pre-Unit and Class 1 demonstrates a critical link to an effective and efficient counting strategy for problem solving.[61] For instance, when a pupil needs to solve an addition problem such as 6 + 3, it is important that they quickly identify that 6 is the bigger number. Pupils who count up from the "bigger number" have learned an effective strategy and also make fewer errors in solving these problems. This task was administered to pupils in grades Pre-Unit through Class 2.

7-10. Addition 1, Addition 2, Subtraction 1, & Subtraction 2

For all four addition and subtraction tasks (Addition 1, Addition 2, Subtraction 1, Subtraction 2), pupils are shown a written representation of the mathematics problem (i.e. an equation), and may also have the problem read aloud to them in order to help them solve the problem.[62] Two levels of addition and subtraction tasks test pupils in Classes 1 through 4 on slightly different levels of difficulty. For Pre-Unit, only Addition Level 1 was administered. Children already have some very basic addition and subtraction concepts before entering formal schooling. For example, children realize that the size of a group of objects grows when more objects are added. (Example: Johnny gets another piece of candy, so Johnny has more candy.) This knowledge is seen in children as young as 3 to 5 years.[63]

A pupil is free to use any method or combination of methods to add and subtract. Possible methods include using fingers, making tick marks on paper with a pen, writing out the problem on the paper with pen, using counters made available to them, or solving problems mentally. If a pupil has not responded or attempted to solve a problem after 10 seconds, the assessor may prompt him or her once, wait 5 seconds, and he or she still does not respond, mark as no response and continue to the next problem. For the addition, subtraction, and word problem tasks below, we include a summary table and an additional graph, which highlights the methods used. Note that each pupil may use more than one method to solve a problem.

Addition 1 was administered to pupils in Pre-Unit and Classes 1 through 4; Addition 2 was administered to pupils in Classes 1 through 4; Subtraction 1 was administered to pupils in Classes 1 through 4; and Subtraction 2 was administered to pupils in Classes 1 through 4.

11. Word Problems

The Word Problems task frames statements to help analyse a child's informal concepts of addition and subtraction. For example, to combine or join two quantities in a word problem is the same as adding or figuring out the sum of two numbers. To assess the strategies children use to solve such problems, the assessor reads the entire word problem to a child before he/she can begin the task. If the child needs a word problem reread, the interviewer rereads it. Each question may be repeated one time at the pupil's request. The child can also use the provided counters in solving the problems.[64] This task was administered to pupils in Classes 1 through 4.

59 Hasbrouck & Tindal, 2006.

60 Tests of oral reading fluency, as measured by timed assessments of correct words per minute, have been shown to have a strong relationship (0.91) with the Reading Comprehension subtest of the Stanford Achievement Test (Fuchs et al., 2001).

61 Clarke et al., 2008.

62 Cooper, Starkey, Blevins, Goth, & Leitner, 1978; Starkey & Cooper, 1980.

63 To test for this knowledge, EGMA uses tasks developed by Jordan, Hanich, and Kaplan (2003).

64 In EGMA, administration of word problems reflects the semantic format provided by Carpenter and Moser (1984) and Riley and Greeno (1988). Based on Carpenter et al.'s (1981) administration of the items, two of each type (joining/separating, part-part-whole, compare) are administered. In Carpenter and Moser's (1984) study, if a child incorrectly answered three out of the first four items or only used the "count all" strategy for the word problems, he/she would not continue with the comparison items. With EGMA, if a child incorrectly answers the first two items or only uses the "count all" strategy for these word problems, he/she does not continue with the following two items. Based on the level of difficulty seen with the comparison items, only joining, separating, and combining are assessed.

A3. Training, Monitoring, and Data Quality

A3.1. Training

DMC conducted a 5-day training for its assessors with the support of Bridge M&E staff. The following is a summary of the activities by day.

Day 1: DMC reviewed pupil characteristics questions and EGRA subtasks, both on paper and on the tablet, to help familiarize assessors with how to use the tablets. This process elicited helpful feedback in terms of edits to make to the instruments. For instance, assessors pointed out that Nursery was frequently referred to as "middle class" and Pre-Unit was frequently referred to as "top class." Changing the words we used would then improve pupil comprehension when asked if they had attended either. DMC then reviewed the correct pronunciations of different letters (Letter Sound Knowledge) and strings of letters (Non-word Reading, Familiar Word Reading).

Day 2: DMC covered the EGMA subtasks and the Pupil Home Life questions at length. We received feedback from attendees and DMC during these sessions as well that was integrated (along with Day 1 edits) into the paper and Tangerine versions of the instruments.

Day 3: DMC provided additional review and coverage of the Letter Sound Knowledge subtask to ensure that assessors were correctly pronouncing letters. DMC used Papaya, a sound pronunciation tool developed by RTI, as part of this session.

Day 4: Training attendees visited 3 different Bridge International Academies to conduct field practice. The selected academies were outside of the evaluation sample. Supervisors oversaw surveyors during assessments to correct any errors and respond to any questions in the sampling methodology. These field visits were instructive, highlighting the importance of ensuring children's comfort, organisation of stimuli and materials, assessor familiarity with subtasks, and consistent application of the sampling methodology for new pupils.

After the field practica, we engaged with DMC in extensive discussions about what difficulties attendees experienced at the site and what additional training or support was needed by assessors.

Day 5: On the last day, assessors continued to work in pairs to practice the EGRA/EGMA subtasks. Additionally, the Bridge M&E Manager reviewed the different types of samples, the sampling strategy, school visit process, and pupil assessment lists with assessors.

A3.2. Monitoring

To ensure that data collection occurred as scheduled and that the DMC team was conducting enough pupil assessments in each of the constituencies participating in the EGRA/EGMA assessment, the M&E team established a number of processes to monitor DMC's daily progress.

Pupil Assessment Lists

The M&E team created pupil assessment lists for each of the different samples. DMC team leaders were responsible for completing these during their visits to each school in the sample groups, and completed lists served as one of the deliverables from DMC. The assessment lists aimed to capture information such as: date of the assessment, beginning and end times of the assessment, pupil's unique randomly-generated Tangerine ID, assessor's name, pupil's full name, and pupil's gender. This information was then used to a) cross-check that data have not been falsified, b) match pupils, and c) identify schools that DMC needed to re-visit. This was particularly important for pupils assessed in 2013, where assessors were to also note the reason for a pupil not being assessed (withdrawn, absent, etc.).

Team Photo

Another deliverable for DMC was one photo of the team of DMC assessors with either the Academy Manager (at Bridge International Academies) or the head teacher (at comparison schools) on the first day of the team's visit to a school. The purpose behind the photo was to serve as verification of DMC's visit to schools (another check to ensure that data were not falsified). The photos will also be used for the cover page of the EGRA/EGMA reports that all participating schools will receive.

Daily Data Downloads

Data was downloaded from Tangerine twice a day during data collection: once at 3.00PM EAT/8.00AM ET and once at 11.00PM EAT/4.00PM ET. These downloads were also backed up. The DMC project manager reviewed this data for his own monitoring purposes to ensure that assessors were collecting and uploading data as scheduled. The M&E team used this data to check on which schools had or had not yet been visited.

Pupil Counts

Using the daily data downloads, the M&E team calculated the number of completed assessments at each school to determine pupil counts. Once these pupil counts had been tallied, short reports were sent to DMC to identify outstanding issues (i.e. missing data in Tangerine – schools scheduled to be visited had no assessments, pupil shortfalls in schools already visited, etc.).

A3.3. Data Quality Checks

In addition to these monitoring systems, there were also re-surveys conducted by the 2 short term field surveyors hired by Bridge. These re-surveys were conducted for pupils in Pre-Unit through Class 4, approximately 1-2 weeks following the original survey conducted by DMC assessors.

There were a total of 222 re-assessments conducted at 68 Bridge academies located in multiple counties. These re-assessments could only be conducted at Bridge as our staff would not be allowed to enter comparison schools.

On average, 3 pupils were assessed at each academy. Across the sample of 222 pupils, we sought to achieve an approximately even distribution across classes and genders based on our sampling process.

The re-assessment data were then uploaded and compared to the original assessments completed by DMC assessors to flag issues of concern that needed to be addressed with DMC (e.g., shorter than average assessment times) and/or ways that subtasks were administered by the field surveyors (e.g., prematurely stopping pupils in a handful of instances).

These comparisons were done on a pupil-by-pupil basis, with each pupil's variables and values compared for both the Re-survey and DMC data. Each pupil comparison also included a summary table that detailed the total number of discrepancies and overall percentages by which the DMC and Re-survey data differed.

Overall, the range of differences was roughly 5% at the lowest end of the spectrum to 30% at the highest, with the average hovering around 15%. These re-survey comparisons likely overestimate the degree to which the two datasets differ for each pupil; this is because the comparisons did not take into account minor spelling differences, extra spaces, or other non-important sources of difference. Therefore, the 15% average is likely an upper bound in terms of discrepancies, and we have no reason to believe that assessment errors were non-random.

A4. Pupil Characteristics

A4.1. Description of Information Collected

Age

Each pupil was asked how old he or she is. Listed ages ranged from 1 to 18, and we also included an "answer extremely unlikely" option for pupils whose stated ages were unlikely to be correct (e.g., a Nursery pupil stating that she was 18 years old).

Gender

Each assessor was asked to record the gender of the pupil being assessed.

ECE Attendance

Depending on the class level of the pupil being tested, pupils were asked if (1) they had attended Baby class, (2) if they had attended Nursery class, and (3) if they had attended Pre-Unit. Pupils in Nursery class, for instance, were not asked if they had attended Pre-Unit.

- *% Attended 1+ Year of ECE:* If pupils attended Baby, Nursery, or Pre-Unit (or any combination of these), they were counted as having attended at least one year of ECE and included in this percentage.
- *If Attended ECE, # of Years*: Recorded as one year per ECE grade level the pupil reported to have attended.

An important caveat is that there may be errors in self-reported information (rather than actual differences between cohorts). Pupils may not remember whether they attended a particular ECE class the further away they are from that time period in their lives. Alternatively, older pupils may be more likely to state that they attended ECE even when they did not, due to their interest in providing a socially desirable response.

Language Spoken at Home

Pupils were asked to specify the languages that they speak at home. We provided a list of languages most commonly spoken in Kenya, including Kiswahili and other mother tongues, as well as English. From these responses, we were able to calculate the following:

- % Speak Their Mother Tongue at Home (Inclusive of Kiswahili)
- % Speak English at Home

Meals

Pupils were asked three different questions about the meals that they had eaten both the day prior to and the day on which they were assessed. We then measured the following:

- % Ate Lunch Day Before
- % Ate Dinner Day Before
- % Ate Breakfast Morning of Assessment

Reading at Home

Pupils were asked if someone read with them at home, and if so, who (e.g., mother, father, grandparent, etc.). We used pupil responses to calculate the following:

- % Someone Reads at Home
- If Someone Reads, % Mother or Father

Homework Assistance

Pupils were asked if someone assisted them with their homework, and if so, whom (e.g., mother, father, grandparent, etc.). Pupil responses were used to calculate the following:

- % Someone Helps with Homework
- If Someone Helps, % Mother or Father

Radio

Pupils were asked a series of questions about radios: (1) did the pupil have a radio in his or her home, (2) did the pupil listen to the radio, and (3) if so, who did the pupil listen to the radio with (e.g., mother, father, uncle, grandparent, and/or friend). Pupil responses were used to calculate the following:

- % Has Radio
- % Listens to Radio
- % Listens to Radio, % with Mother or Father

Television

Pupils were asked a series of questions about televisions, including the following: (1) did the pupil have a TV in his or her home, (2) did the pupil watch TV, and (3) if so, who did the pupil listen to the TV with (e.g., mother, father, uncle, grandparent, and/or friend). Note that a high percentage of pupils have a television in their household: ownership is a sign of prestige and status, particularly in low-income communities, and therefore most families own one.

- % Has TV
- % Watches TV
- If Watches TV, % with Mother or Father

Electricity

Pupils were asked if they have electricity in their homes (this was then used to calculate the variable "% Has Electricity in Home").

A4.2. Baseline Characteristics and Test Scores of Pupils in 2013 and 2014

Table 21. 2013 Pupil Demographics, Education, and Health (Pupils Assessed in 2013 & 2014)

	Pre-Unit			Class 1			Class 2		
	Bridge	Public	B-P	Bridge	Public	B-P	Bridge	Public	B-P
Count	391	438		426	505		469	506	
Demographics									
Mean Age	5.45	6.10	-0.65**	6.72	7.18	-0.46**	7.74	8.23	-0.49**
% Over Target Age	0.12	0.30	-0.18**	0.15	0.32	-0.17**	0.20	0.34	-0.14**
% Female	0.51	0.51	0.00	0.48	0.55	-0.06	0.44	0.52	-0.07*
Early Childhood Education									
% Attended Any ECE	1.00	1.00	0.00	0.92	0.96	-0.04*	0.93	0.96	-0.03*
% Attended Baby	0.53	0.66	-0.13	0.65	0.75	-0.09**	0.73	0.81	-0.08**
% Attended Nursery	0.74	0.80	-0.06	0.78	0.87	-0.1**	0.82	0.90	-0.08**
% Attended Pre-Unit	1.00	1.00	0.00	0.85	0.80	0.04	0.84	0.81	0.03
Meals									
% With No Meals	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average # of Meals	2.94	2.90	0.04	2.93	2.89	0.04*	2.93	2.89	0.04
% Who Had Breakfast	0.98	0.95	0.03*	0.97	0.95	0.02	0.97	0.93	0.04**
% Who Had Lunch	0.98	0.98	0.00	0.98	0.96	0.02	0.98	0.97	0.01
% Who Had Dinner	0.98	0.96	0.01	0.98	0.98	0.00	0.98	0.99	-0.01
Nutrition									
Average DDS	2.34	2.35	-0.01	2.36	2.40	-0.04	2.40	2.41	-0.01
% Had Starch	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
% Had Meat	0.20	0.18	0.02	0.21	0.26	-0.05	0.27	0.28	-0.01
% Had Vitamin	0.01	0.02	-0.01	0.02	0.03	-0.01	0.02	0.01	0.01
% Had Leafy Veg	0.41	0.43	-0.02	0.44	0.39	0.05	0.41	0.44	-0.03
% Had Other Veg	0.19	0.22	-0.03	0.16	0.19	-0.03	0.19	0.17	0.02
% Had Organ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
% Had Eggs	0.03	0.03	0.00	0.05	0.03	0.02	0.02	0.03	0.00
% Had Legumes	0.47	0.44	0.03	0.46	0.48	-0.02	0.46	0.45	0.01
% Had Milk	0.03	0.03	0.01	0.03	0.03	0.01	0.02	0.03	-0.01

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01*

Table 22. 2013 Household Characteristics (Pupils Assessed in 2013 & 2014)

	Pre-Unit			Class 1			Class 2		
	Bridge	Public	B-P	Bridge	Public	B-P	Bridge	Public	B-P
Count	391	438		426	505		469	506	
Assets									
% Has Radio	0.73	0.69	0.04	0.68	0.72	-0.05	0.75	0.73	0.02
% Has TV	0.71	0.55	0.16**	0.70	0.57	0.13**	0.75	0.60	0.14**
% Has Electricity	0.75	0.56	0.2**	0.73	0.57	0.17**	0.78	0.60	0.17**
Language Exposure									
% Speak a Mother Tongue at Home	0.98	0.99	-0.01	0.95	0.94	0.00	0.96	0.98	-0.02
% Speak English at Home	0.07	0.02	0.05**	0.05	0.04	0.02	0.07	0.04	0.04*
% Speak Both at Home	0.05	0.01	0.03**	0.02	0.02	0.00	0.04	0.03	0.01
% Listens to Radio	0.61	0.54	0.07*	0.52	0.58	-0.06	0.59	0.56	0.03
% Watches TV	0.65	0.51	0.14**	0.64	0.50	0.14**	0.68	0.56	0.12**
Parent Involvement									
% Receive Homework Help	0.45	0.39	0.06	0.52	0.46	0.07*	0.44	0.43	0.00
If Given Help, % by Parent	0.55	0.36	0.19**	0.49	0.29	0.2**	0.48	0.31	0.17**
% Are Read To	0.56	0.57	-0.01	0.65	0.58	0.08*	0.58	0.60	-0.02
If Read To, % by Parent	0.49	0.37	0.11*	0.45	0.33	0.12**	0.45	0.30	0.15**
If Listens to Radio, % with Parent	0.79	0.63	0.15**	0.76	0.76	0.00	0.71	0.69	0.02
If Watches TV, % with Parent	0.72	0.64	0.08	0.67	0.64	0.02	0.68	0.64	0.04

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01*

Table 23. 2013 Baseline Test Scores (Pupils Assessed in 2013 & 2014)

	Pre-Unit			Class 1			Class 2		
	Bridge	Public	B-P	Bridge	Public	B-P	Bridge	Public	B-P
Count	391	438		426	505		469	506	
EGRA Zero Scores									
% Zero Listening Comprehension	0.73	0.91	-0.18**	0.47	0.82	-0.36**	--	--	--
% Zero Non-word Reading	0.50	0.59	-0.09**	0.23	0.32	-0.09**	--	--	--
% Zero Familiar Word Reading	0.43	0.58	-0.14**	0.17	0.26	-0.09**	0.03	0.07	-0.03*
% Zero Passage Fluency	0.54	0.68	-0.14**	0.18	0.31	-0.13**	0.04	0.08	-0.04**
% Zero Reading Comprehension	0.92	0.95	-0.03	0.54	0.74	-0.21**	0.22	0.35	-0.13**
EGRA Mean Scores									
Listening Comprehension	0.14	0.04	0.1**	0.32	0.10	0.22**	--	--	--
Non-word Reading	5.71	5.07	0.64	17.03	13.53	3.5**	--	--	--
Familiar Word Reading	4.91	3.34	1.57**	22.57	15.64	6.93**	40.40	35.81	4.58**
Passage Fluency	4.32	3.60	0.72	29.70	19.14	10.56**	55.29	47.07	8.22**
Reading Comprehension	0.02	0.01	0.00	0.15	0.08	0.07**	0.33	0.24	0.09**
EGMA Zero Scores									
% Zero Quantity Discrimination	0.18	0.22	-0.04	--	--	--	--	--	--
% Zero Addition Level I	0.16	0.31	-0.16**	0.03	0.05	-0.03*	0.01	0.01	0.00
% Zero Addition Level II	--	--	--	1.00	1.00	0.00	1.00	1.00	0.00
% Zero Subtraction Level I	--	--	--	0.08	0.13	-0.04*	0.02	0.03	-0.01
% Zero Subtraction Level II	--	--	--	0.39	0.49	-0.1**	0.16	0.25	-0.09**
% Zero Word Problems	--	--	--	0.33	0.45	-0.12**	0.15	0.22	-0.08**
EGMA Mean Scores									
Quantity Discrimination	0.38	0.31	0.07**	--	--	--	--	--	--
Addition Level I	4.33	3.13	1.2**	8.64	7.28	1.36**	11.47	10.91	0.56*
Addition Level II	--	--	--	0.41	0.28	0.13**	0.62	0.52	0.1**
Subtraction Level I	--	--	--	6.54	5.20	1.34**	8.79	8.13	0.66**
Subtraction Level II	--	--	--	0.29	0.20	0.09**	0.47	0.36	0.11**
Word Problems	--	--	--	0.26	0.20	0.06**	0.43	0.36	0.07**

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01*

A4.3. Endline Characteristics and Test Scores of Pupils in 2013 and 2014

Table 24. 2014 Pupil Demographics, Education, and Health (Pupils Assessed in 2013 & 2014)

	Class 1 (Pre-Unit in 2013)			Class 2 (Class 1 in 2013)			Class 3 (Class 2 in 2013)		
	Bridge	Public	B-P	Bridge	Public	B-P	Bridge	Public	B-P
Count	393	445		423	484		471	501	
Demographics									
Mean Age	6.46	7.20	-0.74**	7.72	8.20	-0.48**	8.73	9.24	-0.52**
% Over Target Age	0.12	0.33	-0.21**	0.15	0.32	-0.17**	0.19	0.34	-0.15**
% Female	0.50	0.51	0.00	0.48	0.55	-0.07*	0.44	0.51	-0.07*
Early Childhood Education									
% Attended Any ECE	0.90	0.98	-0.07	0.92	0.97	-0.04**	0.98	0.97	0.01
% Attended Baby	0.64	0.82	-0.18	0.73	0.86	-0.13**	0.74	0.80	-0.06
% Attended Nursery	0.75	0.90	-0.15	0.81	0.91	-0.1**	0.88	0.88	0.00
% Attended Pre-Unit	0.83	0.89	-0.06	0.85	0.86	0.00	0.95	0.95	0.00
Meals									
% With No Meals	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average # of Meals	2.94	2.92	0.02	2.96	2.90	0.06**	2.96	2.93	0.02
% Who Had Breakfast	0.96	0.96	0.00	0.97	0.95	0.01	0.97	0.96	0.00
% Who Had Lunch	0.99	0.99	0.00	1.00	0.96	0.03**	1.00	0.98	0.02*
% Who Had Dinner	0.99	0.98	0.01	1.00	0.99	0.01	0.99	0.99	0.00
Nutrition									
Average DDS	2.48	2.45	0.03	2.53	2.48	0.05	2.58	2.47	0.11*
% Had Starch	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
% Had Meat	0.21	0.25	-0.04	0.24	0.30	-0.06	0.31	0.29	0.02
% Had Vitamin	0.01	0.02	-0.01	0.02	0.02	0.00	0.02	0.01	0.01
% Had Leafy Veg	0.43	0.40	0.02	0.39	0.41	-0.02	0.42	0.46	-0.04
% Had Other Veg	0.28	0.25	0.04	0.28	0.24	0.04	0.25	0.19	0.06*
% Had Organ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
% Had Eggs	0.03	0.02	0.01	0.04	0.05	0.00	0.04	0.04	0.00
% Had Legumes	0.49	0.47	0.03	0.50	0.43	0.07*	0.51	0.46	0.05
% Had Milk	0.03	0.04	-0.01	0.06	0.04	0.02	0.04	0.03	0.01

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01*

Table 25. 2014 Household Characteristics (Pupils Assessed in 2013 & 2014)

	Class 1 (Pre-Unit in 2013)			Class 2 (Class 1 in 2013)			Class 3 (Class 2 in 2013)		
	Bridge	Public	B-P	Bridge	Public	B-P	Bridge	Public	B-P
Count	393	445		423	484		471	501	
Assets									
% Has Radio	0.72	0.73	-0.02	0.70	0.76	-0.06*	0.78	0.75	0.04
% Has TV	0.78	0.57	0.21**	0.79	0.67	0.12**	0.80	0.74	0.07*
% Has Electricity	0.80	0.59	0.21**	0.78	0.64	0.14**	0.80	0.69	0.11**
Language Exposure									
% Speak a Mother Tongue at Home	0.97	0.99	-0.02*	0.98	0.98	0.00	0.97	1.00	-0.03**
% Speak English at Home	0.08	0.03	0.05**	0.08	0.05	0.03	0.11	0.05	0.06**
% Speak Both at Home	0.05	0.03	0.02	0.06	0.04	0.02	0.08	0.04	0.04*
% Listens to Radio	0.60	0.57	0.03	0.56	0.60	-0.04	0.65	0.61	0.04
% Watches TV	0.74	0.54	0.19**	0.74	0.65	0.1**	0.76	0.69	0.07*
Parent Involvement									
% Receive Homework Help	0.64	0.51	0.12**	0.52	0.51	0.02	0.39	0.46	-0.08*
If Given Help, % by Parent	0.55	0.26	0.3**	0.51	0.36	0.14**	0.46	0.31	0.16**
% Are Read To	0.75	0.61	0.14**	0.73	0.62	0.11**	0.64	0.66	-0.01
If Read To, % by Parent	0.57	0.27	0.31**	0.51	0.29	0.21**	0.41	0.31	0.1*
If Listens to Radio, % with Parent	0.77	0.71	0.06	0.72	0.69	0.03	0.66	0.64	0.02
If Watches TV, % with Parent	0.75	0.66	0.09*	0.66	0.62	0.04	0.66	0.61	0.05

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01*

Table 26. 2014 Endline Test Scores (Pupils Assessed in 2013 & 2014)

	Class 1 (Pre-Unit in 2013)			Class 2 (Class 1 in 2013)			Class 3 (Class 2 in 2013)		
	Bridge	Public	B-P	Bridge	Public	B-P	Bridge	Public	B-P
Count	393	445		423	484		471	501	
EGRA Zero Scores									
% Zero Listening Comprehension	0.33	0.71	-0.38**	0.20	0.42	-0.22**	--	--	--
% Zero Non-word Reading	0.14	0.34	-0.2**	0.08	0.12	-0.04*	--	--	--
% Zero Familiar Word Reading	0.08	0.25	-0.17**	0.02	0.07	-0.05**	0.01	0.03	-0.02**
% Zero Passage Fluency	0.11	0.37	-0.26**	0.03	0.10	-0.07**	0.00	0.04	-0.03**
% Zero Reading Comprehension	0.38	0.69	-0.31**	0.12	0.33	-0.22**	0.06	0.16	-0.1**
EGRA Mean Scores									
Listening Comprehension	0.40	0.16	0.24**	0.55	0.35	0.2**	--	--	--
Non-word Reading	17.53	12.41	5.12**	27.43	24.36	3.07**	--	--	--
Familiar Word Reading	23.45	13.78	9.66**	42.30	32.28	10.02**	55.63	45.82	9.8**
Passage Fluency	28.90	17.44	11.46**	62.25	44.37	17.88**	87.19	65.83	21.35**
Reading Comprehension	0.20	0.09	0.12**	0.40	0.23	0.17**	0.47	0.36	0.11**
EGMA Zero Scores									
% Zero Quantity Discrimination	0.01	0.01	0.00	--	--	--	--	--	--
% Zero Addition Level I	0.04	0.07	-0.03*	0.00	0.01	0.00	0.00	0.00	0.00
% Zero Addition Level II	--	--	--	0.09	0.12	-0.03	0.04	0.03	0.01
% Zero Subtraction Level I	--	--	--	0.02	0.03	-0.01	0.01	0.00	0.01
% Zero Subtraction Level II	--	--	--	0.17	0.26	-0.09**	0.10	0.11	-0.01
% Zero Word Problems	--	--	--	0.13	0.23	-0.1**	0.10	0.17	-0.06**
EGMA Mean Scores									
Quantity Discrimination	0.75	0.62	0.13**	--	--	--	--	--	--
Addition Level I	8.36	6.70	1.66**	11.58	10.23	1.35**	13.72	13.63	0.09
Addition Level II	--	--	--	0.59	0.51	0.08**	0.72	0.73	-0.01
Subtraction Level I	--	--	--	8.71	7.19	1.53**	10.23	10.06	0.17
Subtraction Level II	--	--	--	0.46	0.37	0.09**	0.59	0.58	0.01
Word Problems	--	--	--	0.43	0.34	0.09**	0.52	0.47	0.05*

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01*

A5. Panel Fixed Effects

A5.1. Concept

The Panel DiD model with fixed effects is similar to the Panel DiD model including baseline pupil characteristics. However, instead of adding a single variable per baseline characteristic in our regression, applying fixed effects is essentially the same as including one dummy variable per pupil. In other words, we move from parsing out the impact of each characteristic on scores to parsing out the impact of each pupils' characteristics together on scores. This method allows us to hold constant all time-invariant factors associated with each pupil, such as "innate ability", so that any remaining effect is the impact of Bridge. The linear equation for estimating DiD using panel data with fixed effects is:

$$\textbf{\textit{outcome}} = \beta_0 + \beta_1(\textbf{\textit{time}}) + \beta_2(\textbf{\textit{time*treatment}}) + \eta_i + \varepsilon$$

, where

outcome = outcome measure for all units at both points in time;

β_0 = constant term for all units;

β_1 = adjustment for time ; and

β_2 = effect of treatment.

A5.2. Specification

Our resulting panel fixed effects estimator is based on the above linear equation, with the following model specification is as follows:

$$\textit{subtask score} = \beta_0 + \beta_1(\textit{time}) + \beta_2(\textit{time*treatment}) + \eta_i + \varepsilon$$

Just like in the Panel DiD model controlling for baseline characteristics, the treatment effect that we seek to isolate is the "Bridge effect." The fixed effects model, requiring panel data, contains a smaller set of pupils – those baselined in 2013 and re-tested in 2014. We use STATA's -xtreg- command to run our model, clustering at the school level and reporting robust standard errors.

A5.3. Results

We analysed the output generated from our panel fixed effects model in order to compare Bridge pupil performance with that of public school pupils.

When incorporating pupil fixed-effects, Bridge pupils outperform their peers at public schools in four out of five EGRA subtasks. They improved by .14, .22, .28, and .29 standard deviations in non-word reading, familiar word reading, passage fluency, and reading comprehension – above and beyond the growth experienced by their peers. For maths however, the picture is less clear – there is no effect in five of the six subtasks but Bridge underperforms public school pupils in addition level 2.

Notice here that unlike for panel DiD regressions with control variables, we are not able to make any determinations about how specific pupil attributes affect subtask performance levels. This again is due to the fixed effects methodology, which moves away from examining how each characteristic affects test scores, and instead, into holding constant each pupil's characteristics – both observable and unobservable – over time in order to isolate the treatment effect.

Table 27. Panel Fixed Effects Model, Bridge vs. Public Schools

	(1) Listening Comp.	(2) Non-word Reading	(3) Word Reading	(4) Passage Fluency	(5) Reading Comp.	(6) Quantity Disc.	(7) Add Level 1	(8) Add Level 2	(9) Sub Level 1	(10) Sub Level 2	(11) Word Problems
Time (T)	0.521**	0.529**	0.432**	0.459**	0.414**	0.729**	0.459**	0.596**	0.409**	0.558**	0.408**
	(0.0609)	(0.0338)	(0.0221)	(0.0307)	(0.0336)	(0.0569)	(0.0421)	(0.0601)	(0.0622)	(0.0776)	(0.0581)
Bridge	**0.158+**	**0.140****	**0.216****	**0.280****	**0.290****	**0.0952**	**0.0214**	**-0.213***	**-0.00222**	**-0.123**	**0.0332**
Effect (B*T)	**(0.0923)**	**(0.0473)**	**(0.0315)**	**(0.0420)**	**(0.0607)**	**(0.0703)**	**(0.0585)**	**(0.0856)**	**(0.0847)**	**(0.109)**	**(0.0888)**
Count	3518	3517	5472	5472	5148	3517	5472	3813	3813	3812	3812
R-squared	0.222	0.382	0.437	0.351	0.289	0.400	0.220	0.171	0.095	0.152	0.118

*Note: Standard errors in parentheses; + p<0.10, * p<0.05, ** p<0.01. Constant is not shown here, but is included in the regression.*

A6. Guide to Effect Sizes

A6.1. Definition and Calculation

An effect size is the magnitude of difference between two groups – most commonly defined and calculated as the standardized mean difference between two groups. Whereas tests of statistical significance identify whether or not there is a statistically meaningful difference between two groups, effect sizes describe how large that difference is in a more substantive manner.

In their simplest form, effect sizes are calculated as:

$$\text{Effect size} = \frac{\text{Mean(group 1)-Mean (group 2)}}{\text{Pooled Standard deviation}}$$

Group 1 is typically the group that receives an intervention or treatment, whereas *group 2* is the group that does not receive the intervention or treatment. Generally speaking, the unit of comparison is at the student level.[65]

Another way of calculating effect size is by looking at the coefficient on the variable of interest in a regression function.

In the context of education interventions where outcomes are typically measured in test scores, researchers will first standardise scores and turn each test score into a z-score. They do so by first subtracting the sample mean from the raw score for each pupil, and then dividing by the standard deviation of test scores in the sample. Once the scores have been standardised, researchers run a regression. Regression analysis is a statistical tool that allows researchers to identify and characterise relationships between variables in a linear manner.[66] It allows researchers to control for these characteristics depending on the data collected. We can then interpret the coefficient on the treatment variable as the effect size.[67] This is possible because the effect size has already been standardised.

A6.2. Interpretation

Again, standardised effect sizes are used by researchers because they provide clear measures of the magnitude of a difference that exists between two groups. They also allow for different studies to be comparable to one another in a meaningful way.[68] Effect sizes are all measured in standard deviations, rather than study-specific scales, which allows researchers to discuss these in relation to one another using a common metric.

65 The type of standard deviation used to calculate the effect size may differ depending on the design and purpose of the study. (Bloom et al. (2008)

66 Schneider et al. (2010).

67 Just as effect sizes are calculated for statistically significant results, the coefficient on the treatment must be statistically significant for researchers to discuss it as a meaningful effect size.

68 Valentine and Cooper (2003)

Effect size interpretations differ depending on the design of the study, and it is important to note that an effect size does *not* automatically translate to a measure of causal impact:

- **For randomised controlled trials (RCT)** where an intervention is given to one group (treatment) and not to the other (control), the effect size can be interpreted as the direct impact of the intervention. If the effect size is positive, for instance, that would mean that the intervention led to a gain in student learning for the treatment group.
- **For quasi-experimental studies** – which do not randomise – effect sizes can be partly interpreted as the impact of an intervention. The more rigorous the design of the quasi-experimental method, the more the effect translates into impact.
- **For other studies**, the interpretation of the effect size can only be understood as a difference between the two groups. The difference cannot be attributed purely to the treatment at hand.

A6.3. Translation

In the context of the United States, effect sizes are translated into different measures, depending on the design of the study:

- The Center for Research on Education Outcomes' National Charter School Study translates positive effect sizes calculated between charter and non-charter schools into **additional days of student learning**.[69]
- Another interpretation of effect sizes is in the context of what Bloom et al (2008) term as **policy-relevant performance gaps**. For instance, if an intervention improved reading scores by 0.20 for Black American fourth grade students and there was an effect size of -0.83 between Black American and White American fourth grade students, this effect size could be translated as a one fourth reduction in the national Black-White American performance gap.[70]
- In a study of KIPP charter schools conducted by Tuttle et al (2013), effect sizes translate into **identifying movements of student performance from X to Y percentile**.[71]
- The What Works Clearinghouse translates effect sizes into an **"improvement index,"** or what the expected change in percentile rank for an average comparison group student would be had the student received an education intervention.[72]
- Effect sizes in education are also often compared against an "industry benchmark." In the U.S. context, an effect size of 0.25 is used by the What Works Clearinghouse as the benchmark for an effective program with meaningful impacts.[73]

69 Cremata et al. (2013)
70 Bloom et al. (2008)
71 Tuttle et al (2013)
72 What Works Clearinghouse (2008)
73 What Works Clearinghouse (2008). One of the most commonly cited benchmarks for gauging how large an effect size is based on Cohen's paper (1988), whereby he lays out 0.20 as a small effect size, 0.60 as a medium effect size, and 0.80 as a large effect size. Other researchers, however, have argued that these benchmarks are not based on specific empirical cases and that practical measures of effect size should be contextualised appropriately depending on the purpose of the intervention. See Bloom et al. (2008).

A6.4. Calculating Days of Learning

The Stanford CREDO study of charters schools published in 2013 offers the following conversion table:[74]

Table 28. Transformation of Average Learning Gains[75]

Growth (in Standard Deviations)	Days of Learning
0.00	0
0.005	4
0.01	7
0.02	14
0.04	29
0.05	36
0.10	72
0.15	108
0.20	144
0.25	180
0.30	216
0.35	252
0.40	288

However, the key issue with this conversion is that the Stanford CREDO study focuses more on the middle school population, where gains per year measured in effect sizes are typically of a lesser magnitude than in the early grades. Therefore, we turn to an earlier study by Hill et al. to examine the annual reading and maths gains in the U.S. context for grades Kindergarten through Grade 3 – the equivalent to our Pre-Unit through Class 3.

74 Cremata, E. et al (2013). The authors note that the table "[…] shows a sample of standard deviations and their associated days of learning. Positive standard deviations are expressed as additional days of learning; negative standard deviations are associated with fewer days of learning. For those wanting to convert these larger counts into weeks or months: a school week consists of five days, a school month is 20 days, and a quarter or 9-weeks term is typically 45 days. While transforming the statistical results into days of learning provides a more accessible measure, the days of learning are only an estimate and should be used as general guide rather than as empirical transformations."
75 See page 13. Cremata, E. et al (2013).

Table 29. Annual Gains in Standard Deviations by Grade Level

Average Annual Gain in Effect Size From Nationally Normed Tests

	Reading tests		Math tests	
Grade transition	Mean	Margin of error	Mean	Margin of error
Grade K–1	1.52	±0.21	1.14	±0.49
Grade 1–2	0.97	±0.10	1.03	±0.14
Grade 2–3	0.60	±0.10	0.89	±0.16
Grade 3–4	0.36	±0.12	0.52	±0.14
Grade 4–5	0.40	±0.06	0.56	±0.11
Grade 5–6	0.32	±0.11	0.41	±0.08
Grade 6–7	0.23	±0.11	0.30	±0.06
Grade 7–8	0.26	±0.03	0.32	±0.05
Grade 8–9	0.24	±0.10	0.22	±0.10
Grade 9–10	0.19	±0.08	0.25	±0.07
Grade 10–11	0.19	±0.17	0.14	±0.16
Grade 11–12	0.06	±0.11	0.01	±0.14

These transformations are based on data from the US, but serve as our best external source of information to convert gains in effect sizes into days of learning. Such research focused on developing a transformation table for only Kenya or other developing countries has not been conducted to the best of our knowledge.

Therefore, we take the concept underlying the Stanford CREDO study and divided the Bridge effect by the weighted average of the annual gains per grade level (from our EGRA/EGMA study) to get a more precise measure of the additional days of learning.

A7. Pupils Who Stayed vs. Left by School Type

A7.1. Bridge International Academies

Table 30. Differences in Pupil Demographics, Education, and Health:

Attrited vs. Non-Attrited Pupils, Bridge International Academies

	Pre-Unit			Class 1			Class 2			Overall		
	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff
Count	762	555		823	458		873	504		2545	1741	
Demographics												
Mean Age	5.45	5.55	-0.10	6.72	6.84	-0.11	7.74	7.90	-0.16*	7.11	7.35	-0.24**
% Over Target Age	0.12	0.16	-0.04	0.15	0.20	-0.05*	0.20	0.24	-0.04	0.25	0.26	-0.01
% Female	0.51	0.45	0.05	0.48	0.49	0.00	0.44	0.55	-0.1**	0.52	0.46	0.06**
Early Childhood Education												
% Attended Any ECE	1.00	1.00	0.00	0.96	0.95	0.01	0.97	0.96	0.01	0.98	0.96	0.01*
% Attended Baby	0.69	0.71	-0.02	0.77	0.77	0.00	0.84	0.80	0.03	0.77	0.75	0.02
% Attended Nursery	0.83	0.83	0.00	0.88	0.87	0.01	0.90	0.88	0.02	0.87	0.85	0.02
% Attended Pre-Unit	1.00	1.00	0.00	0.85	0.86	-0.01	0.86	0.86	-0.01	0.90	0.90	0.00
Meals												
% With No Meals	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average # of Meals	2.92	2.91	0.01	2.91	2.94	-0.03*	2.92	2.90	0.02	2.92	2.92	0.00
% Who Had Breakfast	0.96	0.96	0.01	0.96	0.96	-0.01	0.95	0.94	0.01	0.96	0.95	0.00
% Who Had Lunch	0.98	0.97	0.00	0.97	0.98	-0.01	0.98	0.99	0.00	0.98	0.98	0.00
% Who Had Dinner	0.97	0.98	0.00	0.98	1.00	-0.01*	0.99	0.98	0.01	0.98	0.98	0.00
Nutrition												
Average DDS	2.43	2.36	0.06	2.46	2.52	-0.06	2.45	2.45	0.01	2.45	2.46	-0.01
% Had Starch	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
% Had Meat	0.22	0.25	-0.03	0.25	0.29	-0.03	0.27	0.26	0.01	0.25	0.26	-0.01
% Had Vitamin	0.01	0.02	0.00	0.02	0.01	0.01	0.01	0.02	-0.01	0.01	0.01	0.00
% Had Leafy Veg	0.42	0.40	0.02	0.43	0.47	-0.04	0.45	0.44	0.01	0.43	0.43	0.01
% Had Other Veg	0.23	0.24	-0.01	0.22	0.19	0.03	0.18	0.20	-0.02	0.21	0.20	0.01
% Had Organ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
% Had Eggs	0.04	0.04	0.00	0.03	0.03	0.00	0.04	0.02	0.01	0.03	0.03	0.00
% Had Legumes	0.47	0.38	0.08**	0.47	0.49	-0.01	0.46	0.45	0.01	0.48	0.47	0.00
% Had Milk	0.04	0.04	0.00	0.03	0.04	0.00	0.04	0.06	-0.02	0.04	0.04	-0.01

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01.*

Table 31. Differences in Baseline Household Characteristics:

Non-Attrited vs. Attrited Pupils, Bridge International Academies

	Pre-Unit			Class 1			Class 2			Overall		
	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff
Count	762	555		823	458		873	504		2545	1741	
Assets												
% Has Radio	0.73	0.73	0.00	0.74	0.71	0.03	0.74	0.74	0.00	0.74	0.72	0.01
% Has TV	0.64	0.64	0.00	0.65	0.67	-0.02	0.72	0.67	0.04	0.67	0.68	0.00
% Has Electricity	0.66	0.66	0.00	0.65	0.69	-0.04	0.72	0.67	0.05	0.68	0.69	0.00
Language Exposure												
% Speak a Mother Tongue at Home	0.98	0.98	-0.01	0.95	0.96	-0.01	0.98	0.98	0.00	0.97	0.98	-0.01
% Speak English at Home	0.04	0.06	-0.02	0.05	0.07	-0.02	0.05	0.04	0.01	0.05	0.05	0.00
% Speak Both at Home	0.02	0.04	-0.02	0.03	0.05	-0.02	0.04	0.03	0.01	0.03	0.04	-0.01
% Listens to Radio	0.60	0.57	0.02	0.60	0.57	0.03	0.57	0.57	0.00	0.59	0.57	0.01
% Watches TV	0.60	0.59	0.00	0.59	0.61	-0.02	0.67	0.61	0.06*	0.63	0.62	0.01
Parent Involvement												
% Receive Homework Help	0.40	0.40	0.00	0.45	0.46	0.00	0.46	0.43	0.03	0.44	0.44	0.00
If Given Help, % by Parent	0.39	0.45	-0.06	0.38	0.40	-0.02	0.39	0.39	0.00	0.39	0.42	-0.03
% Are Read To	0.54	0.51	0.02	0.57	0.57	0.00	0.59	0.59	0.00	0.57	0.56	0.01
If Read To, % by Parent	0.39	0.44	-0.05	0.39	0.35	0.04	0.38	0.36	0.02	0.39	0.39	0.00
If Listens to Radio, % with Parent	0.69	0.73	-0.04	0.76	0.69	0.07*	0.74	0.68	0.06	0.73	0.70	0.03
If Watches TV, % with Parent	0.69	0.65	0.05	0.69	0.66	0.02	0.71	0.64	0.07*	0.70	0.65	0.04*

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01.*

.

Table 32. Differences in Baseline Test Scores

Non-Attrited vs. Attrited Pupils, Bridge International Academies

	Pre-Unit			Class 1			Class 2			Overall		
	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff
Count	762	555		823	458		873	504		2545	1741	
EGRA Zero Scores												
% Zero Listening Comp.	0.87	0.88	-0.01	0.70	0.66	0.04	--	--	--	0.65	0.67	-0.02
% Zero Non-word Reading	0.47	0.56	-0.1**	0.22	0.29	-0.07**	--	--	--	0.25	0.34	-0.1**
% Zero Familiar Word Reading	0.44	0.54	-0.11**	0.17	0.24	-0.07**	0.05	0.10	-0.05**	0.20	0.27	-0.07**
% Zero Passage Fluency	0.53	0.58	-0.06*	0.20	0.28	-0.07**	0.05	0.12	-0.07**	0.24	0.30	-0.05**
% Zero Reading Comp.	0.89	0.89	0.00	0.59	0.59	-0.01	0.24	0.36	-0.12**	0.54	0.56	-0.01
EGRA Mean Scores												
Listening Comprehension	0.07	0.06	0.00	0.19	0.21	-0.02	--	--	--	0.23	0.21	0.02
Non-word Reading	8.49	7.30	1.18*	18.33	17.16	1.17	--	--	--	18.84	15.63	3.22
Familiar Word Reading	6.78	6.30	0.48	22.62	21.16	1.46	40.85	32.68	8.18**	25.13	24.08	1.06
Passage Fluency	7.98	7.65	0.33	28.73	26.65	2.09	56.03	45.91	10.13**	33.51	32.90	0.61
Reading Comprehension	0.03	0.03	0.00	0.14	0.14	-0.01	0.31	0.26	0.06**	0.18	0.17	0.00
EGMA Zero Scores												
% Zero Quantity Disc.	0.17	0.20	-0.03	--	--	--	--	--	--	0.07	0.09	-0.01
% Zero Addition Level I	0.27	0.30	-0.03	0.04	0.08	-0.04**	0.01	0.01	0.00	0.09	0.12	-0.02*
% Zero Addition Level II	--	--	--	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
% Zero Subtraction Level I	--	--	--	0.10	0.12	-0.02	0.02	0.04	-0.02	0.06	0.07	-0.01
% Zero Subtraction Level II	--	--	--	0.43	0.44	-0.01	0.20	0.22	-0.02	0.31	0.28	0.03
% Zero Word Problems	--	--	--	0.39	0.42	-0.03	0.20	0.25	-0.05*	0.28	0.29	-0.01
EGMA Mean Scores												
Quantity Discrimination	0.36	0.33	0.03	--	--	--	--	--	--	0.64	0.62	0.02*
Addition Level I	3.37	3.02	0.35*	7.68	7.20	0.47	11.11	10.57	0.54*	7.78	7.93	-0.15
Addition Level II	--	--	--	0.33	0.32	0.01	0.58	0.58	0.01	0.47	0.51	-0.04**
Subtraction Level I	--	--	--	5.51	5.17	0.34	8.34	7.70	0.64**	7.13	7.24	-0.11
Subtraction Level II	--	--	--	0.24	0.24	0.00	0.42	0.42	0.00	0.34	0.38	-0.03**
Word Problems	--	--	--	0.22	0.21	0.01	0.39	0.35	0.03*	0.32	0.33	-0.01

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01.*

A7.2. Public Schools

Table 33. Differences in Pupil Demographics, Education, and Health:

Attrited vs. Non-Attrited Pupils, Public Schools

	Pre-Unit			Class 1			Class 2			Overall		
	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff
Count	715	661		744	573		836	575		2451	2068	
Demographics												
Mean Age	6.10	6.02	0.07	7.18	7.24	-0.06	8.23	8.24	-0.02	6.77	6.88	-0.11*
% Over Target Age	0.30	0.28	0.02	0.32	0.34	-0.02	0.34	0.35	-0.01	0.15	0.18	-0.04**
% Female	0.51	0.44	0.07	0.55	0.43	0.12**	0.52	0.46	0.06	0.49	0.49	0.00
Early Childhood Education												
% Attended Any ECE	1.00	1.00	0.00	0.94	0.93	0.01	0.95	0.91	0.03*	0.95	0.93	0.03**
% Attended Baby	0.62	0.61	0.01	0.71	0.72	0.00	0.79	0.75	0.05*	0.71	0.68	0.03*
% Attended Nursery	0.80	0.75	0.05*	0.82	0.81	0.02	0.86	0.84	0.02	0.83	0.80	0.02*
% Attended Pre-Unit	1.00	1.00	0.00	0.88	0.87	0.01	0.88	0.81	0.07**	0.91	0.89	0.02*
Meals												
% With No Meals	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average # of Meals	2.94	2.91	0.02	2.93	2.93	0.00	2.94	2.93	0.02	2.94	2.93	0.01
% Who Had Breakfast	0.98	0.96	0.01	0.97	0.96	0.01	0.97	0.96	0.01	0.97	0.96	0.01*
% Who Had Lunch	0.98	0.98	0.00	0.98	0.98	0.00	0.99	0.98	0.01	0.98	0.98	0.00
% Who Had Dinner	0.98	0.97	0.01	0.98	0.99	-0.01	0.98	0.99	0.00	0.98	0.98	0.00
Nutrition												
Average DDS	2.42	2.41	0.02	2.44	2.43	0.02	2.45	2.41	0.04	2.44	2.43	0.01
% Had Starch	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
% Had Meat	0.23	0.21	0.02	0.22	0.22	0.00	0.27	0.23	0.03	0.25	0.23	0.02
% Had Vitamin	0.01	0.01	0.00	0.01	0.01	0.01	0.02	0.02	0.00	0.01	0.01	0.00
% Had Leafy Veg	0.41	0.45	-0.04	0.46	0.44	0.02	0.44	0.43	0.01	0.43	0.44	-0.01
% Had Other Veg	0.21	0.22	-0.01	0.20	0.19	0.01	0.19	0.17	0.02	0.20	0.19	0.01
% Had Organ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
% Had Eggs	0.03	0.05	-0.01	0.04	0.04	0.00	0.03	0.03	0.00	0.04	0.04	-0.01
% Had Legumes	0.49	0.43	0.06*	0.46	0.48	-0.01	0.47	0.48	-0.01	0.47	0.46	0.01
% Had Milk	0.04	0.04	0.01	0.04	0.04	-0.01	0.04	0.05	-0.01	0.04	0.04	0.00

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01.*

Table 34. Differences in Baseline Household Characteristics:

Non-Attrited vs. Attrited Pupils, Public Schools

	Pre-Unit			Class 1			Class 2			Overall		
	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff
Count	715	661		744	573		836	575		2451	2068	
Assets												
% Has Radio	0.76	0.72	0.04	0.72	0.67	0.05	0.75	0.75	0.00	0.74	0.71	0.03*
% Has TV	0.74	0.67	0.07**	0.73	0.70	0.03	0.80	0.73	0.07**	0.76	0.71	0.06**
% Has Electricity	0.77	0.73	0.04	0.76	0.72	0.04	0.82	0.74	0.08**	0.79	0.74	0.05**
Language Exposure												
% Speak a Mother Tongue at Home	0.97	0.98	-0.01	0.96	0.95	0.01	0.97	0.97	0.00	0.96	0.97	0.00
% Speak English at Home	0.07	0.06	0.01	0.06	0.07	-0.01	0.07	0.07	0.00	0.07	0.06	0.01
% Speak Both at Home	0.04	0.04	0.00	0.03	0.04	-0.01	0.05	0.04	0.01	0.04	0.04	0.00
% Listens to Radio	0.64	0.57	0.06*	0.56	0.54	0.03	0.59	0.57	0.01	0.59	0.56	0.04*
% Watches TV	0.68	0.63	0.05	0.68	0.64	0.04	0.75	0.67	0.08**	0.71	0.66	0.05**
Parent Involvement												
% Receive Homework Help	0.43	0.46	-0.03	0.49	0.48	0.01	0.46	0.41	0.06*	0.46	0.45	0.01
If Given Help, % by Parent	0.50	0.50	0.00	0.50	0.48	0.02	0.48	0.50	-0.02	0.50	0.50	0.00
% Are Read To	0.53	0.56	-0.03	0.61	0.58	0.03	0.58	0.57	0.01	0.57	0.57	0.01
If Read To, % by Parent	0.46	0.46	0.00	0.47	0.40	0.07*	0.47	0.48	-0.01	0.47	0.46	0.02
If Listens to Radio, % with Parent	0.77	0.69	0.08*	0.77	0.70	0.07*	0.75	0.73	0.02	0.76	0.71	0.05**
If Watches TV, % with Parent	0.73	0.67	0.06*	0.70	0.67	0.03	0.73	0.70	0.03	0.71	0.68	0.04*

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01.*

Table 35. Differences in Baseline Test Scores

Non-Attrited vs. Attrited Pupils, Public Schools

	Pre-Unit			Class 1			Class 2			Overall		
	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff	Stayed	Left	Diff
Count	715	661		744	573		836	575		2451	2068	
EGRA Zero Scores												
% Zero Listening Comp.	0.77	0.79	-0.02	0.48	0.51	-0.03	--	--	--	0.51	0.57	-0.07**
% Zero Non-word Reading	0.41	0.52	-0.11**	0.16	0.26	-0.1**	--	--	--	0.22	0.35	-0.13**
% Zero Familiar Word Reading	0.35	0.48	-0.13**	0.11	0.20	-0.09**	0.03	0.07	-0.05**	0.17	0.27	-0.1**
% Zero Passage Fluency	0.44	0.55	-0.1**	0.12	0.22	-0.11**	0.03	0.09	-0.05**	0.18	0.27	-0.09**
% Zero Reading Comp.	0.87	0.88	-0.01	0.45	0.54	-0.08**	0.17	0.28	-0.11**	0.46	0.53	-0.08**
EGRA Mean Scores												
Listening Comprehension	0.13	0.12	0.01	0.32	0.28	0.04*	--	--	--	0.32	0.27	0.06
Non-word Reading	9.06	7.55	1.51**	20.84	17.71	3.13**	--	--	--	19.80	15.71	4.09
Familiar Word Reading	7.87	6.59	1.28*	27.32	22.60	4.72**	43.65	37.42	6.23**	27.95	24.28	3.67**
Passage Fluency	8.66	7.52	1.14	35.80	28.05	7.75**	61.04	51.86	9.18**	39.04	34.99	4.05**
Reading Comprehension	0.04	0.03	0.00	0.19	0.15	0.04**	0.37	0.31	0.06**	0.22	0.19	0.03**
EGMA Zero Scores												
% Zero Quantity Disc.	0.15	0.17	-0.03	--	--	--	--	--	--	0.06	0.08	-0.02**
% Zero Addition Level I	0.18	0.22	-0.05*	0.02	0.05	-0.03**	0.01	0.01	-0.01	0.06	0.09	-0.03**
% Zero Addition Level II	--	--	--	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
% Zero Subtraction Level I	--	--	--	0.07	0.09	-0.02	0.02	0.02	0.00	0.04	0.05	-0.01
% Zero Subtraction Level II	--	--	--	0.37	0.40	-0.03	0.15	0.19	-0.04	0.24	0.27	-0.02
% Zero Word Problems	--	--	--	0.32	0.35	-0.03	0.16	0.19	-0.03	0.22	0.24	-0.02
EGMA Mean Scores												
Quantity Discrimination	0.41	0.38	0.03	--	--	--	--	--	--	0.69	0.65	0.04**
Addition Level I	4.04	3.65	0.38*	8.50	8.08	0.42	11.43	11.09	0.34	8.45	8.26	0.19
Addition Level II	--	--	--	0.40	0.37	0.03	0.65	0.62	0.03	0.54	0.54	0.01
Subtraction Level I	--	--	--	6.31	6.08	0.23	8.72	8.45	0.27	7.80	7.99	-0.20
Subtraction Level II	--	--	--	0.29	0.27	0.02	0.49	0.46	0.03	0.41	0.40	0.01
Word Problems	--	--	--	0.26	0.24	0.02	0.43	0.40	0.03	0.36	0.36	0.00

*Note: * indicates p-value below 0.05, ** indicates p-value below 0.01.*

A8. Differential Attrition Regressions

A8.1. Differential Attrition Probits

Table 36. Likelihood of Attrition, Bridge vs. Public Schools

	Early Grade Reading Assessment					Early Grade Math Assessment					
	(1)	(2)	(3)	(4)	(5)	(8)	(9)	(10)	(11)	(12)	(13)
	Listening Comp.	Non-Word Reading	Word Reading	Passage Fluency	Reading Comp.	Quantity Disc.	Add Level I	Add Level II	Sub Level I	Sub Level 2	Word Prob.
Bridge	1.322*	1.448*	1.092*	1.031*	0.982+	1.539*	1.109*	0.759	1.021	0.760	0.819
	(0.660)	(0.630)	(0.515)	(0.525)	(0.537)	(0.631)	(0.523)	(0.651)	(0.644)	(0.657)	(0.649)
Baseline Score	0.0257	-0.139*	-0.102*	-0.0807	-0.00332	-0.141**	-0.115**	-0.00190	-0.0941*	-0.00321	-0.0328
	(0.0674)	(0.0684)	(0.0519)	(0.0607)	(0.0636)	(0.0458)	(0.0429)	(0.0415)	(0.0411)	(0.0525)	(0.0530)
Baseline Score*Bridge	-0.0830	0.0240	-0.0244	-0.0575	-0.0759	0.0306	-0.0326	-0.0524	0.0687	-0.0578	-0.0102
	(0.0758)	(0.0835)	(0.0621)	(0.0692)	(0.0726)	(0.0588)	(0.0564)	(0.0535)	(0.0576)	(0.0654)	(0.0690)
Age	-0.0127	-0.00624	0.0123	0.00889	-0.00104	0.0138	0.0241	0.0181	0.0289	0.0181	0.0210
	(0.0313)	(0.0303)	(0.0233)	(0.0228)	(0.0234)	(0.0289)	(0.0224)	(0.0278)	(0.0286)	(0.0288)	(0.0290)
Female	-0.249**	-0.235**	-0.202**	-0.206**	-0.218**	-0.228**	-0.207**	-0.212**	-0.204**	-0.212**	-0.211**
	(0.0710)	(0.0743)	(0.0515)	(0.0522)	(0.0522)	(0.0723)	(0.0498)	(0.0534)	(0.0518)	(0.0533)	(0.0530)
Years of ECD	0.0377	0.0407	0.0108	0.00820	0.00147	0.0432	0.00828	-0.0369	-0.0295	-0.0370	-0.0353
	(0.0426)	(0.0428)	(0.0382)	(0.0387)	(0.0375)	(0.0432)	(0.0383)	(0.0459)	(0.0463)	(0.0460)	(0.0467)
Has Radio	-0.0547	-0.0532	-0.0914	-0.0938	-0.0964	-0.0472	-0.0926	-0.139	-0.150	-0.139	-0.140
	(0.128)	(0.126)	(0.113)	(0.114)	(0.114)	(0.128)	(0.115)	(0.117)	(0.121)	(0.116)	(0.118)
Has TV	0.0543	0.0622	0.154	0.153	0.151	0.0807	0.151	0.221*	0.223+	0.221*	0.218+
	(0.134)	(0.135)	(0.103)	(0.104)	(0.104)	(0.134)	(0.103)	(0.112)	(0.114)	(0.112)	(0.113)
Has Electricity	0.0994	0.114	0.130	0.123	0.111	0.105	0.117	0.146	0.151	0.146	0.146
	(0.101)	(0.0998)	(0.0846)	(0.0838)	(0.0838)	(0.0982)	(0.0810)	(0.0957)	(0.0948)	(0.0954)	(0.0955)
Mother Tongue	0.406*	0.400*	0.399**	0.396**	0.385**	0.426*	0.396**	0.331*	0.353*	0.331*	0.335*
	(0.176)	(0.174)	(0.147)	(0.147)	(0.146)	(0.179)	(0.146)	(0.161)	(0.160)	(0.160)	(0.160)
English	0.347*	0.395*	0.231	0.226	0.214	0.396*	0.231	0.199	0.214	0.198	0.207
	(0.172)	(0.166)	(0.145)	(0.145)	(0.149)	(0.176)	(0.144)	(0.138)	(0.137)	(0.136)	(0.141)
Listens to Radio	-0.000187	-0.00811	0.0355	0.0369	0.0421	-0.0190	0.0295	0.0583	0.0531	0.0585	0.0582
	(0.126)	(0.128)	(0.103)	(0.102)	(0.101)	(0.125)	(0.103)	(0.104)	(0.105)	(0.104)	(0.105)

Watches TV	-0.00482	-0.0130	-0.0281	-0.0271	-0.0276	-0.0252	-0.0254	-0.0999	-0.0921	-0.0998	-0.0919
	(0.154)	(0.154)	(0.120)	(0.120)	(0.118)	(0.154)	(0.119)	(0.131)	(0.133)	(0.131)	(0.131)
Read to at Home	-0.0795	-0.0753	-0.0664	-0.0664	-0.0759	-0.0879	-0.0623	-0.0262	-0.0240	-0.0262	-0.0273
	(0.0743)	(0.0743)	(0.0582)	(0.0580)	(0.0596)	(0.0745)	(0.0592)	(0.0751)	(0.0747)	(0.0755)	(0.0760)
Receives HW Help	0.0490	0.0472	0.0317	0.0339	0.0399	0.0584	0.0323	0.0265	0.0263	0.0265	0.0276
	(0.101)	(0.0999)	(0.0717)	(0.0718)	(0.0716)	(0.0996)	(0.0722)	(0.0802)	(0.0806)	(0.0802)	(0.0805)
Meal Count	0.0158	0.0271	-0.0150	-0.0163	-0.0226	0.0194	-0.0105	0.00509	0.0216	0.00517	0.00690
	(0.102)	(0.101)	(0.0753)	(0.0754)	(0.0757)	(0.104)	(0.0743)	(0.0944)	(0.0909)	(0.0945)	(0.0935)
Diet Diversity Score	0.00557	0.0156	0.0225	0.0190	0.0176	0.0131	0.0177	0.0578	0.0567	0.0579	0.0577
	(0.0541)	(0.0528)	(0.0464)	(0.0469)	(0.0464)	(0.0519)	(0.0459)	(0.0640)	(0.0631)	(0.0628)	(0.0630)
Age*Bridge	0.0303	0.0309	0.0283	0.0330	0.0319	0.0226	0.0245	0.0420	0.0276	0.0414	0.0375
	(0.0390)	(0.0389)	(0.0290)	(0.0286)	(0.0294)	(0.0365)	(0.0288)	(0.0392)	(0.0401)	(0.0401)	(0.0395)
Female*Bridge	0.207*	0.199+	0.278**	0.283**	0.283**	0.191+	0.277**	0.366**	0.356**	0.363**	0.364**
	(0.102)	(0.105)	(0.0755)	(0.0758)	(0.0751)	(0.104)	(0.0728)	(0.0799)	(0.0790)	(0.0796)	(0.0805)
Years of ECD*Bridge	-0.0380	-0.0350	-0.0493	-0.0481	-0.0431	-0.0432	-0.0468	-0.00544	-0.0122	-0.00445	-0.00643
	(0.0568)	(0.0565)	(0.0471)	(0.0475)	(0.0467)	(0.0573)	(0.0472)	(0.0562)	(0.0565)	(0.0564)	(0.0572)
Has Radio*Bridge	0.0114	0.00744	0.0846	0.0904	0.0883	0.0136	0.0815	0.122	0.126	0.120	0.118
	(0.173)	(0.173)	(0.141)	(0.141)	(0.141)	(0.173)	(0.142)	(0.161)	(0.163)	(0.160)	(0.161)
Has TV*Bridge	-0.426+	-0.405+	-0.411*	-0.410*	-0.418*	-0.428+	-0.404*	-0.288	-0.299	-0.291	-0.289
	(0.222)	(0.226)	(0.180)	(0.179)	(0.178)	(0.224)	(0.180)	(0.205)	(0.208)	(0.206)	(0.206)
Has Electricity*Bridge	-0.158	-0.176	-0.199+	-0.193	-0.185	-0.168	-0.195+	-0.277+	-0.283+	-0.279+	-0.278+
	(0.138)	(0.138)	(0.119)	(0.119)	(0.119)	(0.137)	(0.117)	(0.145)	(0.145)	(0.145)	(0.145)
Mother Tongue*Bridge	-0.487*	-0.506*	-0.449*	-0.444*	-0.431*	-0.529*	-0.434*	-0.298	-0.328	-0.295	-0.305
	(0.230)	(0.230)	(0.209)	(0.209)	(0.209)	(0.233)	(0.212)	(0.238)	(0.238)	(0.237)	(0.239)
English*Bridge	-0.526*	-0.577**	-0.354+	-0.344+	-0.329+	-0.591**	-0.351+	-0.193	-0.222	-0.201	-0.205
	(0.217)	(0.213)	(0.181)	(0.182)	(0.185)	(0.220)	(0.184)	(0.201)	(0.201)	(0.199)	(0.203)
Listens to Radio*Bridge	-0.0235	-0.0144	-0.103	-0.106	-0.107	-0.0157	-0.0946	-0.122	-0.117	-0.126	-0.124
	(0.171)	(0.174)	(0.134)	(0.133)	(0.133)	(0.172)	(0.134)	(0.150)	(0.151)	(0.150)	(0.151)
Watches TV*Bridge	0.287	0.274	0.248	0.251	0.252	0.289	0.228	0.152	0.150	0.163	0.152
	(0.219)	(0.221)	(0.180)	(0.180)	(0.177)	(0.219)	(0.180)	(0.213)	(0.214)	(0.213)	(0.214)

Read to at Home*Bridge	0.0178	0.00834	0.0207	0.0196	0.0276	0.0394	0.0240	-0.0254	-0.0225	-0.0242	-0.0224
	(0.107)	(0.106)	(0.0884)	(0.0885)	(0.0891)	(0.106)	(0.0881)	(0.110)	(0.110)	(0.111)	(0.110)
Receives HW Help*Bridge	0.0224	0.0200	-0.0117	-0.0127	-0.0145	-0.00150	-0.0216	-0.0904	-0.0919	-0.0895	-0.0914
	(0.115)	(0.115)	(0.0885)	(0.0884)	(0.0882)	(0.114)	(0.0889)	(0.104)	(0.105)	(0.105)	(0.104)
Meal Count*Bridge	-0.228	-0.256	-0.116	-0.119	-0.107	-0.236	-0.121	-0.0692	-0.0876	-0.0682	-0.0699
	(0.161)	(0.159)	(0.116)	(0.117)	(0.118)	(0.159)	(0.117)	(0.147)	(0.144)	(0.146)	(0.145)
Diet Diversity Score*Bridge	0.0486	0.0450	0.00629	0.00931	0.00504	0.0441	0.0117	-0.0534	-0.0510	-0.0543	-0.0519
	(0.0751)	(0.0754)	(0.0640)	(0.0644)	(0.0636)	(0.0744)	(0.0636)	(0.0788)	(0.0775)	(0.0778)	(0.0777)
Count	2815	2815	4408	4408	4405	2815	4408	3071	3071	3070	3071
Pseudo R-squared	0.038	0.041	0.045	0.045	0.043	0.044	0.046	0.052	0.053	0.052	0.052
Joint Test (X2): Characteristics	44.43	51.88	61.68	60.88	60.94	47.36	60.04	52.28	55.22	51.24	53.46
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Joint Test (X2): Demographic	6.313	5.589	18.34	18.10	18.07	5.530	18.33	21.79	23.39	21.59	21.57
P-value	0.177	0.232	0.001	0.001	0.001	0.237	0.001	0.000	0.000	0.000	0.000
Joint Test (X2): Assets	7.853	5.955	11.97	12.75	13.13	6.296	12.14	8.602	9.130	7.728	7.352
P-value	0.097	0.203	0.018	0.013	0.011	0.178	0.016	0.072	0.058	0.102	0.118
Joint Test (X2): Language	9.857	10.63	6.714	7.281	7.987	10.68	7.127	3.452	3.600	2.939	2.088
P-value	0.020	0.014	0.082	0.064	0.046	0.014	0.068	0.327	0.308	0.401	0.554
Joint Test (X2): Learning	3.631	1.971	2.576	3.508	4.228	2.480	2.768	2.940	4.014	2.598	2.318
P-value	0.604	0.853	0.765	0.622	0.517	0.779	0.736	0.709	0.547	0.762	0.804
Joint Test (X2): Health	3.626	2.838	0.996	1.429	1.723	2.752	1.282	1.376	2.865	1.330	0.740
P-value	0.305	0.417	0.802	0.699	0.632	0.431	0.733	0.711	0.413	0.722	0.864

*Note: Standard errors in parentheses. Significance: + p<0.10, * p<0.05, ** p<0.01. Non-interacted control variables such as county dummies were also included but not shown here. Pupils in Kajiado, Kiambu, Kilifi, Nyandarua, and Taita-Taveta are less likely to attrite than pupils in Nairobi. Pupils in Nyamira are more likely to attrite. Bridge pupils in Kajiado, Kiambu, and Kilifi were more likely to attrite than those in Nairobi relative to their public school peers, and Bridge pupils in Nyamira and Mombasa were less likely to attrite.*

A8.2. BGLW Pooling Test

Table 37. BGLW Test Results, Bridge vs. Public Schools

	Early Grade Reading Assessment					Early Grade Math Assessment					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Listening Comp.	Non-word	Word Reading	Passage Fluency	Reading Comp.	Quantity Disc.	Add Level I	Add Level II	Sub Level I	Sub Level I	Word Prob.
Attrited, 2013 to 2014	-0.555	-0.196	0.665+	0.480	0.394	-0.945	-0.374	-1.036+	-0.585	0.0325	-0.734
	(-1.46)	(-0.33)	(1.67)	(1.29)	(0.95)	(-1.15)	(-1.23)	(-1.68)	(-1.12)	(0.05)	(-0.92)
Bridge*Attrite	-0.173	-0.693	-0.712	-1.069	-0.685	0.373	0.118	1.604+	0.510	0.121	1.040
	(-0.25)	(-0.75)	(-1.11)	(-1.46)	(-1.00)	(0.35)	(0.23)	(1.85)	(0.58)	(0.13)	(1.09)
Age*Bridge*Attrite	-0.0266	-0.0656+	-0.0191	-0.0241	-0.0493+	-0.0556	0.00749	-0.0206	0.0141	-0.0249	-0.0421
	(-0.82)	(-1.79)	(-0.68)	(-0.92)	(-1.88)	(-1.14)	(0.33)	(-0.44)	(0.35)	(-0.53)	(-0.91)
Female*Bridge*Attrite	0.106	0.137	0.0246	0.0327	-0.0440	0.0584	-0.0453	-0.164	-0.115	-0.0298	0.0366
	(0.89)	(1.42)	(0.35)	(0.50)	(-0.60)	(0.43)	(-0.76)	(-1.50)	(-1.05)	(-0.27)	(0.36)
Years of ECD*Bridge*Attrite	0.00674	0.115+	0.0123	0.0186	0.00228	0.0128	0.0522	-0.0342	0.134*	0.130*	0.0266
	(0.11)	(1.79)	(0.30)	(0.43)	(0.05)	(0.15)	(1.25)	(-0.45)	(2.28)	(2.38)	(0.41)
Has Radio*Bridge*Attrite	0.285	0.138	-0.00417	0.0155	0.134	0.168	0.0693	-0.387+	0.114	-0.131	-0.198
	(1.64)	(0.81)	(-0.03)	(0.11)	(1.03)	(0.84)	(0.53)	(-1.68)	(0.56)	(-0.55)	(-1.01)
Has TV*Bridge*Attrite	0.124	0.384	0.348*	0.223	0.0818	-0.00165	0.0877	0.213	-0.114	0.106	0.518+
	(0.50)	(1.56)	(2.32)	(1.53)	(0.45)	(-0.01)	(0.44)	(0.67)	(-0.40)	(0.38)	(1.67)
Has Electricity*Bridge*Attrite	0.0563	0.0151	-0.0893	-0.170+	-0.109	0.198	0.0465	0.0368	-0.0772	0.247	0.218
	(0.46)	(0.11)	(-0.90)	(-1.93)	(-1.18)	(0.82)	(0.49)	(0.18)	(-0.48)	(1.33)	(1.51)
Mother Tongue*Bridge*Attrite	-0.721*	0.0120	0.173	0.191	0.227	0.0350	-0.391+	-0.907*	-0.937**	-0.491	-0.127
	(-2.28)	(0.03)	(0.52)	(0.60)	(0.57)	(0.07)	(-1.81)	(-2.60)	(-3.97)	(-1.58)	(-0.29)
English*Bridge*Attrite	-0.550	0.0385	-0.199	-0.129	0.127	-0.0165	-0.164	-0.574+	-0.408+	0.368	-0.261
	(-1.58)	(0.13)	(-0.93)	(-0.59)	(0.49)	(-0.04)	(-0.86)	(-1.98)	(-1.78)	(1.25)	(-0.59)
Listens to Radio*Bridge*Attrite	-0.0863	-0.0451	0.0831	0.0302	-0.130	0.0383	-0.00459	0.443*	0.185	0.0256	0.122
	(-0.52)	(-0.29)	(0.71)	(0.27)	(-1.14)	(0.22)	(-0.04)	(2.01)	(1.04)	(0.11)	(0.67)
Watches TV*Bridge*Attrite	-0.351	-0.525*	-0.327*	-0.144	-0.0823	-0.0456	-0.102	-0.311	0.0284	-0.328	-0.698*
	(-1.45)	(-2.35)	(-2.06)	(-0.95)	(-0.42)	(-0.16)	(-0.53)	(-0.95)	(0.09)	(-1.23)	(-2.39)
Read to at Home*Bridge*Attrite	-0.117	0.180	0.141	0.0968	0.0715	0.269+	0.133	0.0505	0.140	0.291*	0.203
	(-0.87)	(1.44)	(1.64)	(1.19)	(0.74)	(1.66)	(1.45)	(0.31)	(1.04)	(2.00)	(1.31)
Receives HW Help*Bridge*Attrite	-0.0339	-0.145	-0.213*	-0.215*	-0.247**	-0.338*	-0.357**	-0.273+	-0.308*	-0.179	-0.199
	(-0.30)	(-1.21)	(-2.48)	(-2.62)	(-2.77)	(-2.30)	(-4.84)	(-1.75)	(-2.59)	(-1.28)	(-1.24)
Meal Count*Bridge*Attrite	0.361*	0.299	0.193	0.305	0.267	-0.136	0.0801	-0.135	0.124	0.0685	-0.259
	(2.00)	(1.31)	(1.31)	(1.59)	(1.53)	(-0.54)	(0.59)	(-0.68)	(0.55)	(0.32)	(-1.38)
Diet Diversity Score*Bridge*Attrite	-0.0433	-0.0800	0.0402	0.0540	0.0131	-0.00938	-0.0524	0.0722	-0.108	-0.0602	0.0210
	(-0.47)	(-1.00)	(0.94)	(1.27)	(0.22)	(-0.09)	(-0.98)	(0.79)	(-1.18)	(-0.59)	(0.22)
Count	2815	2815	4408	4408	4405	2815	4408	3071	3071	3070	3071
R-squared	0.196	0.266	0.440	0.443	0.344	0.348	0.440	0.213	0.211	0.186	0.188
Joint Test (F)	2.702	1.807	1.679	1.847	2.001	0.842	2.849	1.386	2.507	1.526	0.970
P-value	0.00253	0.0505	0.0749	0.0441	0.0268	0.623	0.00149	0.178	0.00488	0.119	0.490

*Note: Standard errors in parentheses. Significance: + p<0.10, * p<0.05, ** p<0.01. Non-interacted control variables such as baseline characteristics and county dummies, along with these variables interacted with the attrite dummy, were also included but not shown here. The county dummies interacted with Bridge and the attrite dummy are also not included. The joint test was conducted for all baseline characteristics interacted with Bridge and the attrite dummy (with the exception of county dummies).*

The Bridge Effect

A Comparison of Early Grade Learning Gains in English and Maths

Impact Evaluation Report