Cross-Age Tutoring: Experimental Evidence from Kenya

Mauricio Romero* Lisa Chen† Noriko Magari†

Abstract

There is an increasing wealth of evidence showing that teaching appropriate to the student's learning level can improve learning outcomes in low-income countries. Cross-age tutoring, where older students tutor younger students, is an inexpensive alternative for providing personalized instruction to younger students at the cost of the older student's time. We present the results from a large RCT in Kenya, in which schools are randomly assigned to implement either an English or a math tutoring program. Students in grades 3-7 tutor students in grades 1-2 and preschool. We find that tutoring in math, relative to tutoring in English, has a small positive effect (0.06 SD, p-value of 0.073) on math test scores. These results do not hold true for English tutoring, however: relative to math tutoring, it has no positive effect on English test scores (we can rule out an effect of 0.077 SD with 95% confidence). We show that there is considerable heterogeneity according to the student's baseline learning level: The effect is largest for students in the middle of the ability distribution (0.144 SD, p-value of 0.005), while the point estimates are almost zero for students with either very low or very high baseline learning levels. Finally, we show that tutors are neither harmed by nor benefit from the program.

1 Introduction

Over the past three decades access to primary school has dramatically increased in low- and middle-income counties (United Nations, 2015). However, the quality of education remains poor despite increases in enrollment.[1] This is especially worrisome as evidence suggests that the quality of education, not the quantity, is what matters for growth (Hanushek & Kimko, 2000; Hanushek & Wößmann, 2007).

*University of California - San Diego.
†Bridge International Academies

[1]For example, despite enrollment rates of over 90%, less than 50% of children in Argentina, Colombia, Morocco, Uganda, Namibia, and Malawi attain "minimum literacy standards" (*Global Monitoring Report*, 2007). Estimates from Mexico and Brazil estimate that more than 50% of children lack minimal competency in mathematics (Filmer, Hasan, & Pritchett, 2006). Jones, Schipper, Ruto, and Rajani (2014) show that the majority of children in grade 3 across East Africa cannot recognize a single word in their medium of instruction.

Interventions that tailor teaching to student learning levels are consistently signaled by the literature as having the largest effects on learning levels across different settings (for three recent reviews of the literature see Glewwe and Muralidharan (2016); Evans and Popova (2016); Snilstveit et al. (2016)). However, teachers often lack the time (or incentives) to give each child personalized instruction tailored to their needs and providing schools with extra teachers to do so is expensive. Cross-age tutoring, where older students tutor younger students, is an inexpensive alternative to providing personalized instruction to younger students in that it substitutes a trained instructor (the teacher) with an untrained one (the older student) at the cost of the older student's time. However, to the extent that tutoring can also provide benefits to tutors (e.g., mastering knowledge and increasing social skills), cross-age tutoring could result in an overall welfare improvement. In this article, we present the results from a large RCT (over 180 schools, 20,000 tutees, and 20,000 tutors) in Kenya, in which schools are randomly selected to implement a cross-age tutoring program in either English or math.

A noteworthy feature of this RCT is that all schools in our sample implement a tutoring program (i.e., there is no "pure" control group that receives no tutoring at all). The random assignment determines whether a school implements an English or a math tutoring program. Therefore, all of our results should be interpreted as the impact of tutoring in math relative to tutoring in English (or vice versa). Although the lack of a "pure" control group may give the impression that our estimates are difficult to interpret, in this RCT we know exactly what the counterfactual for time use is across groups. An alternative would have been to provide tutoring to some schools and no tutoring to others. However, since time is finite, we would either need to control how tutors and tutees use the time allocated for tutoring in control schools or let schools/students choose how to use this time. Is not clear that either case would lead to a better counterfactual.

Cross-age tutoring has existed since at least 95 CE (Quintilianus & Halm, 1869) and is now widely used across the world. In the typical setup, the tutor is a few years older than the tutee and works with him or her on specific problems in a particular subject. Cross-age tutoring is often thought to have positive effects for both tutors and tutees in terms of academic achievement and social-emotional outcomes (Cohen, Kulik, & Kulik, 1982). However, the evidence on the subject is mixed, based on small-scale experiments or observational data and mostly from developed countries. An early review of the literature showed that cross-age tutoring had a positive effect (both in terms of academic performance and attitudes) on both tutees and tutors (Cohen et al., 1982). However, a more recent review looking only at randomized control trials comes to the conclusion that cross-age tutoring in math has non-significant effects on math test scores and cross-age tutoring in "reading" has a small (statistically significant) positive effect on reading (Shenderovich, Thurston, & Miller, 2015).

In our setting tutoring took place each school day of the 2016 academic year. At the end of every day, older students tutored younger students in either English or math for 40 minutes. Tutors were five grades above tutees. In some schools the tutoring focusesd on math, while in others it focused in English. Whether math or English tutoring took place was randomized across schools. Therefore, within a schools all grades participated in either math or English tutoring. Section 2.2 has details on the math and English tutoring interventions.

Our results show that cross-age tutoring in math, relative to tutoring in English, has a small positive effect (0.06 SD, p-value of 0.073) on math test scores. These results do not hold true for English tutoring, however: relative to math tutoring, it has no positive effect on English test scores (we can rule out an effect of 0.077 SD with 95% confidence). We show that there is considerable heterogeneity according to the student's baseline learning level. Specifically, the effect of math tutoring, relative to English tutoring, on math test scores is largest for students in the middle of the ability distribution (0.144 SD, p-value of 0.005), while the point estimate is almost zero for students with either very low or very high baseline learning levels. This is consistent with: a) tutors not being able to help students who are advanced learners and need an instructor with a high level of expertise to guide them through more advanced material; and b) tutors not being able to help tutees lagging behind grade level competencies who may need more specialized instruction to catch up.

In addition, we find that there is no heterogeneity by tutees' gender or age. Similarly, we show that there is no heterogeneity by school characteristics (pupil-teacher ratio, class size, or tutor-tutee ratio). Since we do not have data on tutor/tutee matches and teacher had discretion on how to match tutees to tutors, we can only show heterogeneity by the average characteristics of possible tutors for a specific tutee; we find that there is no heterogeneity in tutors' average age, gender or proficiency level (baseline test scores).

Taken together, these results imply that in this setting math tutoring is more effective than English tutoring in raising test scores (in the subject of tutoring), and that although tutors are limited in their ability to help certain students, they can effectively increase test scores for students in the middle of the distribution of baseline learning levels.

Our results speak directly to three strands in the literature. First, they relate to the literature that seeks to understand the underlying production function for cognitive achievement (Todd & Wolpin, 2003) and studies the impact of different education policies and programs (see (Glewwe & Muralidharan, 2016;

Evans & Popova, 2016; Snilstveit et al., 2016) for recent reviews of the literature). We present evidence on a novel approach to improve the amount of personalized instruction at low cost. Second, they speak to the literature on peer effects and their effect on learning (Sacerdote, 2001; Zimmerman, 2003; Munley, Garvey, & McConnell, 2010). We explore how the quality of different tutors (at the school level) affects the outcomes of the program. Finally, we communicate with the literature on peer-learning programs. Only two of the studies reviewed by Shenderovich et al. (2015) had other elementary school students (as opposed to adults, community volunteers, or university students) as tutors and both tutoring programs focus on reading. None of the interventions in those studies were implemented in a low- or middle-income country. To the best of our knowledge, ours is the first RCT implemented on cross-age tutoring in which tutors are students in the same school as tutees. Furthermore, it is the first study of this kind in a low-income country.

The article is organized as follows. The next section presents an overview of the experimental design, the context, the intervention and the sampling strategy. Section 3 presents the empirical strategy and the results of the analysis. Finally, Section 4 concludes.

2 Experimental Design

2.1 Context

Despite high net enrollment rates in primary schools (∼95% in 2013), the quality of education in Kenya is low: Children often fail to attain proficiency in early grade reading and numeracy. The annual nationwide learning assessments carried out by Twaweza/Uwezo consistently show that only half of grade 3 students can read a simple story at a grade 2 level in English (the national language and the language of instruction) or successfully demonstrate grade 2 numerical skills (Jones et al., 2014).

Bold, Kimenyi, and Sandefur (2013) argue that the abolition of fees for primary schools in 2003 led to a decline in the quality ("or at least perceived quality") of public schools and in response the demand (and supply) of private primary education increased dramatically. According to World Bank statistics, the proportion of students enrolled in private primary schools more than doubled from 4.5% in 2004 to over 10.5% in 2009.

Kenya is not the only country that has seen a surge in private school enrollment. Recently several chains of for-profit, low-cost private schools have emerged around the world. These chains leverage technology to deliver lessons and to manage teachers more effectively (Mbiti, 2016). In this article, we work with a large low-cost private school provider, Bridge International Academies (Bridge), in which schools within their network are randomly selected to implement either a math or an English tutoring program. Bridge opened its first school in a Nairobi slum in January 2009. By November 2014, it had opened nearly 400 schools across Kenya and had enrolled over 100,000 students.[2]

Bridge takes advantage of economies of scale in school management, teacher training, and teaching guides to deliver education at a low cost.[3] English is the language of instruction in all Bridge schools, which are located across East Africa, but mainly in Kenya. The company relies heavily on technology to maintain a constant.[4]

From a research standpoint, an advantage of working with Bridge data is that all students take the same tests across all schools, and Bridge collects data on students' performance to provide teachers with feedback. Students are tested six times per academic year. Each academic year has three terms, and each term has a midterm and an endterm exam. Additionally, at the beginning of the academic year students in primary grades (Standards 1 - 6) take a diagnostic exam. This means that RCTs to study the effectiveness of different approaches to improve learning can be implemented relatively easily with no additional cost for data collection (often the most expensive part of an RCT). This is the first of such RCTs implemented across schools in the Bridge network.

2.2 Intervention

The intervention took place every school day during the 2016 academic year. At the end of every school day, older students tutored younger students in either English or math for 40 minutes (3:35-4:15 pm). Tutors are five grades above tutees (See Table 1 for details). In some schools the tutoring focused on math, while in others it focused in English. Whether math or English tutoring took place was randomized across schools. Therefore, within a schools all grades participated in either math or English tutoring. Section 2.2.1 has details on the math tutoring intervention, while Section 2.2.2 provides details on the English tutoring intervention.

[2]See `http://www.bridgeinternationalacademies.com/company/history/`

[3]For example, in each Bridge academy, unlike in "mom-and-pop" low-cost private schools, there is only one employee involved in management. This is because the vast majority of non-instructional activities that the Bridge "Academy Manager" would normally have to deal with (billing, payments, expense management, payroll processing, and more) are automated and centralized. Similarly, Bridge hires experts to develop comprehensive teacher guidelines and training programs, which are then used in all of their schools. Schools charge on average a monthly fee of USD 6 and cater to families living on USD 2 a day per person or less.

[4]Bridge follows the 8-4-4 curriculum framework mandated by the national government, but provides detailed teacher guides for each lesson that are used by teachers across the network. These guides are created by Bridge headquarter offices in both Boston, MA and Nairobi, KE, and are then streamed to teachers' personal tablets. Teachers use tablets to upload students' information (e.g., test scores) to Bridge country headquarter offices as well.

Table 1: Tutors and Tutees

Tutors	Tutees
Grade 3 →	Baby Class (BC)
Grade 4 →	Nursery (NU)
Grade 5 →	Preunit (PU)
Grade 6 →	Grade 1
Grade 7 →	Grade 2

The main objective of the math (English) tutoring program was to raise math (English) achievement in tutees (Baby-Grade 2 pupils). A secondary objective was to develop communication and leadership skills in tutors (Grade 3-Grade 7 pupils) and build a school community through sibling-like relationships between tutees and tutors.

Tutors were given guides with problems and activities to do with tutees each day. Teachers lead the mentoring session, deliver the "tutor" guide, and choose how to pair tutees with tutors. Teacher could also vary the matching between tutees to tutors each day. During the first two weeks of the academic year 2016 the mentoring sessions consisted of "mentor training". During this mentor training, teacher instructed mentors to keep pupils focused and use the "ask-tell-show-repeat" procedure to correct pupil work. "Ask-tell-show-repeat" is a four-step process in which tutees are asked to do a problem again if they answer incorrectly; they then receive verbal instructions on the correct solution if the mistake is repeated; they are then shown the correct solution if they make a mistake again; and finally, the pupil is asked to repeat the problem one last time. The idea was to provide a simple structure for mentor-pupil interaction.

For math, minor changes were introduced during the last quarter of the school year. Specifically, brief instructions for mentors replace the teacher demonstration. This was done to shift the focus of the mentoring session from the teacher to the mentoring pairs. "Tutor guides" then instruct tutors to provide immediate feedback to tutees, telling them whether the answer was correct or incorrect as soon as they answer a question. A simplified version of the "ask-tell-show-repeat" correction method took its place: the "ask-show-repeat" method. Instead of first verbally instructing the pupil in how to obtain a correct solution in case of a repeat mistake, the mentor now goes straight to showing them the correct solution. Finally, teachers were instructed to "check-respond-leave" with mentors exclusively, thus empowering mentors to take responsibility for their pupils' performance (See Table 2 for details.

For English, minor changes were introduced during the last two quarters of the school year. Most of the changes varied how much time was allocated to different activities (See Table 3 for details).

2.2.1 Math tutoring

Table 2: Math tutoring intervention

	Term 1 and Term 2	Term 3
Timing	3:35 - 4:15pm.	3:35 - 4:15pm.
C1/C2	Intro: 3 min Teacher demo: 5 min Mentoring: 30 min Guide with 18 problems 1 topic	Intro: 3 min Mentoring 1: 22 min Mentoring 2: 15 min Guide with 60 problems 2 topics
PU	Intro: 3 min Warm-up Exercise: 10 min Teacher demo: 5 min Mentoring: 15 min Guide wtih 10 problems 1 topic	Intro: 3 min Mentoring 1: 22 min Mentoring 2: 15 min Guide with 56 problems 2 topics
BC/NU	Introduction: 3 min Counting with mentors: 7 min Rhyme: 3 min ID numbers with mentors: 7 min ID frames with mentors: 7 min Rhyme: 3 min ID shapes with mentors: 8 min Closing: 2 min	Introduction: 3 min Counting with mentors: 7 min Rhyme: 3 min Writing numbers with mentors: 7 min Drawing frames with mentors: 7 min Rhyme: 3 min Drawing shapes with mentors: 8 min Closing: 2 min
Tutor duties	Keep pupils focused Use ask-tell-show-repeat	Correct tutee after every two problems Use ask-show-repeat
Teacher duties	Do teacher demo Circulate	Check-respond-leave with mentors only

2.2.2 English tutoring

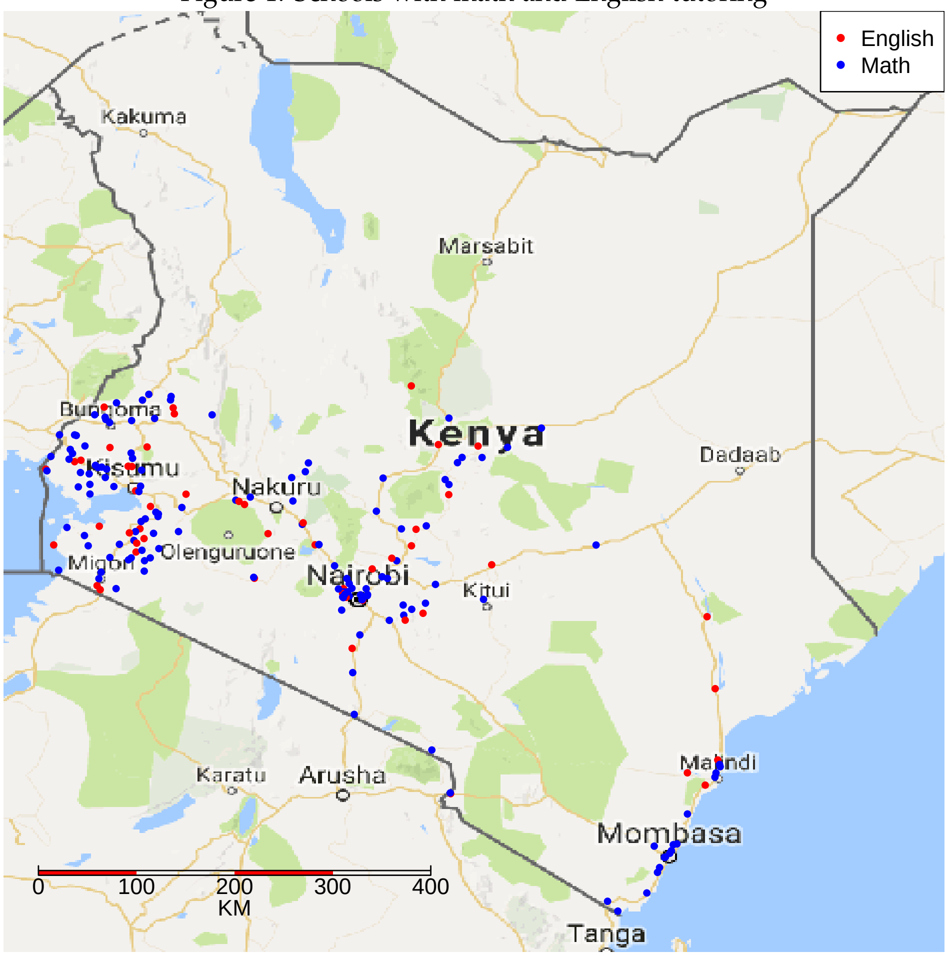
Table 3: English tutoring intervention

	T1	T2	T3
Timing	3:35 - 4:15pm.	3:35 - 4:15pm.	3:35 - 4:15pm.
C1/C2	Introduction: 2 min Dialogue practice: 5 min Mentoring instructions: 3 min Words: 5 min Reading: 15 min Writing: 9 min	Introduction: 2 min Dialogue practice: 5 min Words: 8 min Writing: 15 min Reading: 9 min	Introduction: 3 min Words: 10 min Writing: 10 min Reading: 15 min Closing: 2 min
PU	Introduction: 3 min Dialogue practice: 5 min Practice book: 7 min Sight words: 5 min Reading: 15 min	Introduction: 3 min Dialogue practice: 5 min Sight words: 12 min Reading: 15 min	Introduction: 2 min Words: 8 min Reading: 15 min Sight words: 15 min
BC/NU	Introduction & song: 5 min Practice set: 7 min Finding words: 4 min Rhyme: 3 min Finding letters: 5 min Letter sound chant: 2 min Dialogue practice: 5 min Closing: 2 min	Introduction & song: 5 min Words: 11 min Rhyme: 3 min Finding letters: 5 min Letter sound chant: 2 min Dialogue practice: 5 min Closing: 2 min	Introduction & song: 5 min Words: 11 min Rhyme: 3 min Finding rhyme words: 7 min Letter sound chant: 2 min Finding letters: 5 min Dialogue practice: 5 min Closing: 2 min
Mentor duties	Keep pupils focused Use ask-tell-show-repeat Correction method	Keep pupils focused Use ask-tell-show-repeat Correction method	Keep pupils focused Use ask-tell-show-repeat Correction method
Teacher duties	Circulate	Circulate	Circulate

2.3 Sampling

Bridge has a network of over 400 schools across Kenya. However, only 187 schools were eligible to participate in the trial.[5]. Randomization was stratified at the province level and by average baseline test scores at each academy. Estimations take into account the randomization design by including the appropriate fixed effects (Bruhn & McKenzie, 2009). Figure 1 shows the distribution of schools across the country. Math tutoring took place in 137 academies, while English tutoring took place in 50 academies.

Figure 1: Schools with math and English tutoring



2.4 Data and summary statistics

As mentioned before, students are tested six times per academic year. Each academic year has three terms, and each term has a midterm and an endterm exam. Additionally, at the beginning of the academic year

[5]Schools where a pilot of the program was tested during the 2015 academic year were excluded.

students in primary grades (Standards 1 - 6) take a diagnostic exam. Table 4 shows the dates of each exam. Two exams (T3ET15 and T1DG16) were taken by students before tutoring began, and six exams were taken after. Since students in Preunit, Nursery and Baby-Class are not tested at the beginning of 2016 (T1DG16), we use both T1DG16 and T3ET15 as our baseline test scores. Note that for students in Baby Class (BC) we have no baseline test scores.

All students across schools in Bridge's network take the same exam, making test scores for students in different schools comparable. However, the exams are not vertically linked (i.e., there are no overlapping questions across exams), and therefore we standardized test scores in each term (to obtain mean zero and standard deviation of 1 in English tutoring schools).

Table 4: Learning assessments

Year	**Term**	**Exam**	**Dates**	**Code**
2015	3	Endterm	11/10/2015 - 11/12/2015	T3ET15
2016	1	Diagnostic	1/13/2016 - 1/14/2016	T1DG16
2016	1	Midterm	2/16/2016 - 2/18/2016	T1MT16
2016	1	Endterm	4/5/2016 - 4/7/2016	T1ET16
2016	2	Midterm	6/14/2016 - 6/16/2016	T2MT16
2016	2	Endterm	8/9/2016 - 8/11/2016	T2ET16
2016	3	Midterm	9/26/2016 - 9/27/2016	T3MT16
2016	3	Endterm	10/25/2016 - 10/27/2016	T3ET16

Tables 5-7 show summary statistics of schools, tutees and tutors in schools with math and English tutoring. Table 5 shows that schools randomly assigned to math tutoring are similar to those assigned to English tutoring: They were inaugurated around the same time (in operation for two years by January 1, 2016), and have similar teacher salaries and pupil-teacher ratios (PTR) of 22 students per teacher. Pupils (Table 6) in English and math tutoring schools are similar across all characteristics.[6] Tutors (Table 7) are also similar across English and math tutoring schools. On average pupils are 6.5 years old and tutors are 4.5 years older than their tutees.

[6]Except for Science and Social Sciences in T1DG16 where students in English tutoring schools seem to be doing better. However, neither of these subjects is the focus of tutoring, and moreover when correcting for multiple-hypothesis testing these differences are no longer significant.

Table 5: School characteristics in English and math tutoring schools

	English Tutoring	Math Tutoring	Difference	Difference (F.E)
Days since launch date (as of Jan/01/2016)	672.960	693.310	20.347	-16.257
	(406.417)	(405.017)	(66.887)	(46.838)
Monthly teacher wage of 11250 KSH	0.060	0.110	0.049	0.014
	(0.240)	(0.313)	(0.043)	(0.026)
Monthly teacher wage of 10400 KSH	0.180	0.100	-0.078	-0.073
	(0.388)	(0.304)	(0.061)	(0.061)
Monthly teacher wage of 7970 KSH	0.760	0.790	0.028	0.058
	(0.431)	(0.410)	(0.070)	(0.065)
Teachers	7.440	7.530	0.093	0.077
	(0.541)	(0.619)	(0.093)	(0.092)
Pupils	167.760	167.180	-0.585	-2.554
	(75.793)	(84.627)	(12.894)	(11.451)
PTR	22.240	21.980	-0.257	-0.478
	(9.363)	(10.367)	(1.589)	(1.401)

Days since launch date indicates the number of days that have passed since the schools opened, as of January 1, 2016. Bridge has three teacher wage categories. “Monthly teacher wage” shows the proportion of schools within each wage schedule. Teachers is the number of teachers at the school, and pupils is the enrollment across all grades for the school at the beginning of the school year. PTR is the pupil-teacher ratio.

Each row presents the mean for schools that receive English tutoring (Column 1), schools that receive math tutoring (Column 2), the difference between the two (Column 3), and the difference taking into account the randomization design (Column 4). In the first two columns the standard deviation is shown in parentheses, while in the third and fourth columns the standard error of the difference is in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Pupil's characteristics: ECE, Grade 1 and Grade 2

Panel A: Invariant characteristics				
	English Tutoring	Math Tutoring	Difference	Difference (F.E)
Age	6.510	6.430	-0.081	-0.017
	(1.662)	(1.631)	(0.055)	(0.036)
Male	0.520	0.520	0.002	0.000
	(0.500)	(0.500)	(0.010)	(0.010)
Age entered Bridge	5.470	5.430	-0.038	0.023
	(1.686)	(1.663)	(0.075)	(0.069)
Panel B: T3ET15				
	English Tutoring	Math Tutoring	Difference	Difference (F.E)
English (Reading)	0.000	-0.010	-0.013	-0.058
	(1.000)	(1.021)	(0.074)	(0.071)
English (Writing)	0.000	-0.040	-0.038	-0.064
	(0.999)	(1.014)	(0.064)	(0.058)
Swahili (Reading)	0.000	-0.020	-0.025	-0.064
	(1.000)	(1.020)	(0.083)	(0.082)
Swahili (Writing)	0.000	-0.070	-0.072	-0.113
	(1.000)	(1.102)	(0.111)	(0.090)
Development	0.000	0.060	0.055	0.022
	(1.000)	(1.043)	(0.101)	(0.093)
Environmental	0.000	-0.040	-0.039	-0.043
	(0.999)	(1.066)	(0.065)	(0.059)
Math	0.000	0.040	0.041	0.011
	(0.999)	(0.974)	(0.056)	(0.052)
Science	0.000	-0.070	-0.069	-0.098
	(1.000)	(1.009)	(0.089)	(0.081)
S.S.	0.000	-0.020	-0.018	-0.055
	(1.000)	(1.030)	(0.093)	(0.084)
Panel C: T1DG16				
	English Tutoring	Math Tutoring	Difference	Difference (F.E)
English	0.000	-0.050	-0.047	-0.077
	(1.000)	(1.061)	(0.082)	(0.078)
Math	0.000	-0.060	-0.056	-0.086
	(1.000)	(1.060)	(0.085)	(0.087)
Science	0.000	-0.160	-0.164*	-0.180**
	(1.000)	(1.064)	(0.089)	(0.080)
S.S.	0.000	-0.130	-0.128	-0.147**
	(1.000)	(1.053)	(0.077)	(0.073)
Swahili	0.000	0.030	0.026	-0.001
	(1.000)	(1.053)	(0.078)	(0.076)

Math, Language (English), Swahili, Science, and S.S. (Social Sciences) represent the standardized test scores (mean zero and standard deviation 1 in English tutoring schools).

Each row presents the mean for schools that receive English tutoring (Column 1), schools that receive math tutoring (Column 2), the difference between the two (Column 3), and the difference taking into account the randomization design (Column 4). In the first two columns the standard deviation is shown in parentheses, while in the third and fourth columns the standard error of the difference is in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Pupil's characteristics: Grade 3 - Grade 7

Panel A: Invariant characteristics				
	English Tutoring	Math Tutoring	Difference	Difference (F.E)
Age	11.030	11.060	0.028	0.022
	(1.983)	(2.005)	(0.095)	(0.061)
Male	0.500	0.520	0.025^{***}	0.027^{***}
	(0.500)	(0.499)	(0.008)	(0.008)
Age entered Bridge	9.740	9.790	0.052	0.047
	(2.275)	(2.309)	(0.135)	(0.096)
Panel B: T3ET15				
	English Tutoring	Math Tutoring	Difference	Difference (F.E)
English (Reading)	0.000	0.070	0.070	0.047
	(0.999)	(1.038)	(0.051)	(0.046)
English (Writing)	0.000	0.070	0.069	0.034
	(0.999)	(0.967)	(0.054)	(0.045)
Swahili (Reading)	0.000	0.050	0.055	0.053
	(0.999)	(1.042)	(0.056)	(0.046)
Swahili (Writing)	0.000	0.140	0.138^{*}	0.114^{*}
	(0.999)	(0.941)	(0.081)	(0.059)
Math	0.000	0.050	0.047	0.027
	(0.999)	(1.009)	(0.063)	(0.048)
Science	0.000	0.060	0.060	0.026
	(0.999)	(1.010)	(0.052)	(0.044)
S.S.	0.000	0.070	0.070	0.034
	(0.999)	(0.967)	(0.056)	(0.049)
Panel C: T1DG16				
	English Tutoring	Math Tutoring	Difference	Difference (F.E)
English	0.000	0.040	0.041	0.027
	(0.999)	(1.030)	(0.050)	(0.048)
Math	0.000	0.050	0.046	0.030
	(0.999)	(0.999)	(0.050)	(0.044)
Science	0.000	0.030	0.027	0.011
	(0.999)	(1.051)	(0.044)	(0.040)
S.S.	0.000	0.050	0.051	0.043
	(0.999)	(1.027)	(0.047)	(0.044)
Swahili	0.000	-0.010	-0.010	-0.010
	(0.999)	(1.017)	(0.065)	(0.041)

Math, Language (English), Swahili, Science, and S.S. (Social Sciences) represent the standardized test scores (mean zero and standard deviation 1 in English tutoring schools).

Each row presents the mean for schools that receive English tutoring (Column 1), schools that receive math tutoring (Column 2), the difference between the two (Column 3), and the difference taking into account the randomization design (Column 4). In the first two columns the standard deviation is shown in parentheses, while in the third and fourth columns the standard error of the difference is in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A common feature in our data is that we have an unbalanced panel, where very few students have test score data for all periods in all grades. This is the result of software/network failures in which the teacher enters the data but it fails to upload to servers at Bridge HQ.[7] Table 8 shows the fraction of students tested each time. As can be seen, more than 30% of the data is missing (and often more than 40%). In particular, the endterm exam in the in the second period (T2ET16) is missing over 60% of the test scores for math. There was a glitch in the software that prevented more than 25% of the academies from entering test-score data for T2ET16. Since this is noisy data, we remove it from our sample in the main text, but we provide robustness checks that include the data in Appendix C.

Table 8: Missing data

	T1MT16	T1ET16	T2MT16	T2ET16	T3MT16	T3ET16	Total
Math	0.690	0.546	0.690	0.390	0.569	0.532	0.569
	(0.463)	(0.498)	(0.463)	(0.488)	(0.495)	(0.499)	(0.495)
English (writing)	0.679	0.532	0.689	0.460	0.563	0.517	0.573
	(0.467)	(0.499)	(0.463)	(0.498)	(0.496)	(0.500)	(0.495)
English (reading)	0.679	0.524	0.688	0.438	0.552	0.512	0.565
	(0.467)	(0.499)	(0.463)	(0.496)	(0.497)	(0.500)	(0.496)
Observations	199350						

This table shows the fraction of students with scores for math, English (reading), and English (writing) in each test.

In Table 9 we show that whether the data for a particular student is missing is orthogonal to whether that student is receiving math or English tutoring. Since attrition is high in any given period (over 30%) we do not perform Lee (2009) bounds as these are too wide to be informative.[8] Additionally, since a large number of students do not have baseline test scores we input scores for those students and add a dummy variable to all our regressions for whether the baseline test score was inputted.

[7]In addition, students may be absent from school on the day of the test. However, in most cases if test score data is missing for a student, it is also missing for their entire grade.

[8]Note that the T2ET16 testing rates are different across math and English tutoring (see Figure B.2). However, we believe that this difference is the result of chance.

Table 9: Differential attrition between treatment and control students

	Tested
Math tutoring	-0.011
	(0.018)
Mean English	0.40
N. of obs.	84095
Number of schools	187

Clustered standard errors, by school, in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table shows the differential attrition between students in math tutoring schools compared to students in English tutoring schools. The estimation data set does not include T2ET16 data. See Table C.1 for an estimation data set that includes T2ET16 data.

3 Results

3.1 Treatment effects

In order to estimate the effect of tutoring on test scores we use the following specification:

$$Y_{isgd,t} = \alpha_0 + \beta T_s + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t} \quad (1)$$

where $Y_{isgd,t}$ is the test score of student i in grade g at school s located in district d at time t (and $Y_{isgd,t=0}$ is his test score before treatment). γ_d is a set of district and strata fixed effects, γ_t are time fixed effects, and γ_g are grade fixed effects. X_i is a set of student time-invariant characteristics (month of birth and gender), and X_s are school characteristics at baseline (pupil-teacher ratio, monthly school fees and teachers' wages). T_s is a dummy that indicates whether the student is in a school with a math tutoring program (if not, he is in a school with English tutoring). Standard errors are clustered at the school level. β is the coefficient of interest here and estimates the effect of math tutoring on test scores compared to English tutoring. Notice that this specification assumes that the treatment effect (β) is time invariant and grade invariant (in Section 3.2 we relax these assumptions).

3.1.1 Tutees

Table 10 shows the effect of math tutoring (relative to English tutoring) on math and English for tutees. Math tutoring has a small positive effect on math test scores of 0.6 SD. However, English tutoring (or the lack of it) has no effect on English test scores–we can rule out an effect bigger than $0.077SD$ with a confidence of 95%. Neither math nor English tutoring seem to have no effect on Swahili (or other subjects XXXX NEED to add other subjects table XXXXXXXXX).

Table 10: Effect on tutees' test scores

	Math	English	Swahili
Math tutoring	0.062*	-0.0040	0.036
	(0.035)	(0.035)	(0.047)
N. of obs.	50461	48241	32755
Number of schools	187	187	186

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does not include T2ET16 data. See Table C.2 for a version of this table that includes T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

3.1.2 Tutors

Table 11 shows that the tutoring program had no effect on tutors. We can rule out an effect greater than 0.09 SD with a confidence of 95% in math (for the math tutoring program). Similarly, we can rule out an effect greater than 0.06 SD with a confidence of 95% in English (for the English tutoring program).

Table 11: Effect on tutees' test scores

	Math	English	Swahili
Math tutoring	0.029	-0.019	-0.020
	(0.031)	(0.035)	(0.036)
N. of obs.	48741	46938	46512
Number of schools	187	187	187

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does not include T2ET16 data. See Table C.3 for a version of this table that includes T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

3.2 Heterogeneity

In this section we test for heterogeneous treatment effects in tutees.[9] Overall, we find that the math tutoring program is most effective after the first period (except for T2ET16, the period with a high attrition rate and therefore unreliable results). We also find that math tutoring is most effective for students in the middle of the ability distribution at baseline. We do not find any heterogeneity by grade, age, gender, average tutor characteristics (age, gender, baseline test scores), or average school characteristics (pupil-teacher ratio, school size, or tutor-tutee ratio).

3.2.1 Periods

In order to estimate the effect of tutoring on test scores across time we use the following specification

$$Y_{isgd,t} = \alpha_0 + \sum_{\tau=1}^{6} \beta_\tau T_s \times 1_{t=\tau} + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t}, \quad (2)$$

where β_1 measures the treatment effect in period T1MT15, β_2 measures the treatment effect in T1ET15, and so on, until β_6 which measures the treatment effect in period T3ET15. Figure 2 shows the treatment effect in each period for each subject. As can be seen, the treatment effect for math (of math tutoring relative to English tutoring) increases after the first marking period (except for T2ET16, the period with a high attrition rate). On the other hand, math tutoring, relative to English tutoring, does not seem to have a negative effect on English test scores, with point estimates close to zero after the first marking period.

[9]Results for tutors are available upon request.

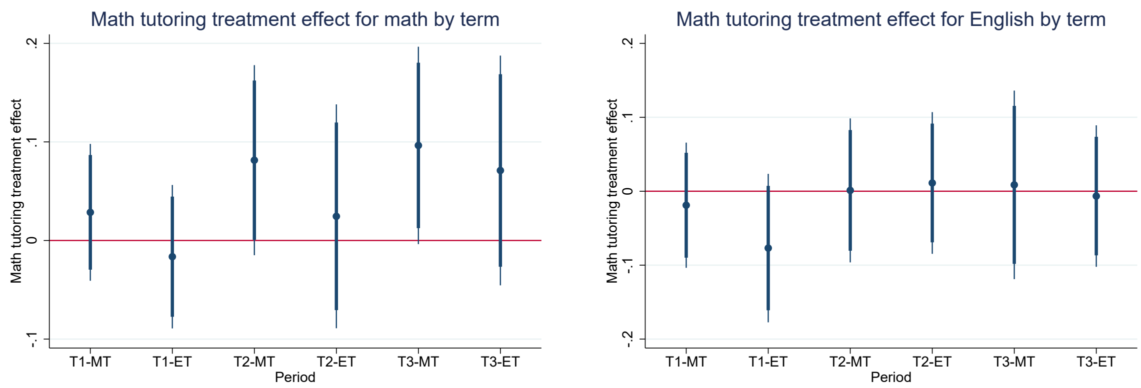


Figure 2: Evolution of the treatment effect of math tutoring, relative to English tutoring, on math (left panel) and English (right panel) test scores. Bars represent 90% and 95% confidence intervals (thick lines and thin lines, respectively).

3.2.2 Grade

In order to estimate the effect of tutoring on test scores across time we use the following specification

$$Y_{isgd,t} = \alpha_0 + \sum_{\tau=1}^{5} \beta_\tau T_s \times 1_{t=\tau} + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t}, \tag{3}$$

where β_1 measures the treatment effect for BC, β_2 for NU, β_3 for PU, β_4 for STD 1 and β_5 for STD 2. Although the point estimate for STD 2 is the largest in math, there does not seem to be a systematic pattern in which oldest students benefit more than younger ones from math tutoring, and we cannot reject the hypothesis that the effect is the same across grades. Similarly, there seems to be no systematic pattern in the effect on English test scores.

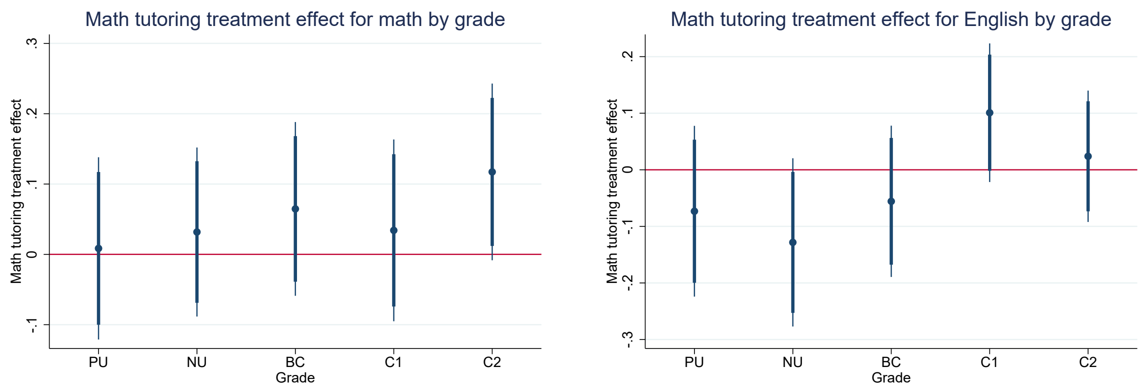


Figure 3: Treatment effect of math tutoring, relative to English tutoring, in math (left panel) and English (right panel) test scores by grade. Bars represent 90% and 95% confidence intervals (thick lines and thin lines, respectively).

3.2.3 Tutee characteristics

In order to estimate the effect of tutoring on test scores across tutee characteristics we use the following specification:

$$Y_{isgd,t} = \alpha_0 + \beta_1 T_s + \beta_2 T_s \times c_i + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t}, \quad (4)$$

where c_i denotes the characteristics along which we wish to measure heterogeneity and β_2 allows us to test whether there is any differential treatment effect. Panel A show that there is evidence of heterogeneity by tutee's age. Specifically, older students benefit more from tutoring. It is important to note that in this context the age distribution in each grade has wide tails and they often overlap (see Figure B.1 in Appendix B). Panel B shows that there is no heterogeneity by gender. This stands in contrast to several education interventions in which girls benefit more.[10] As Panel C shows, there is no heterogeneity by the age at which students enrol at Bridge.

[10] For example, Anderson (2008) reviews several early childhood interventions with larger effects for girls and Chetty, Friedman, and Rockoff (2014) reports larger impacts of teacher quality on girls than boys.

Table 12: Heterogeneity by tutees' characteristics

	Math	English	Swahili
Panel A: Age			
Math tutoring	0.055	-0.015	-0.00033
	(0.034)	(0.035)	(0.051)
Date of birth	-0.042**	0.031	0.091**
	(0.020)	(0.024)	(0.037)
Math tutoring × Date of birth	-0.037*	-0.053**	-0.077*
	(0.022)	(0.026)	(0.045)
Observations	50821	48598	33069
Adjusted R^2	0.222	0.293	0.335
Panel B: Gender			
Math tutoring	0.053	-0.017	0.036
	(0.033)	(0.035)	(0.046)
Male	0.0030	-0.029**	-0.021
	(0.013)	(0.013)	(0.014)
Math tutoring × Male	-0.013	-0.0028	-0.013
	(0.014)	(0.015)	(0.016)
Observations	50935	48705	33119
Adjusted R^2	0.220	0.293	0.335
Panel C: Bridge enrollment age			
Math tutoring	0.055	-0.016	0.019
	(0.033)	(0.034)	(0.047)
AgeFirstBridge	0.013	-0.065***	-0.067**
	(0.020)	(0.021)	(0.030)
Math tutoring × AgeFirstBridge	0.039*	0.038	0.050
	(0.021)	(0.024)	(0.033)
Observations	50821	48598	33069
Adjusted R^2	0.221	0.294	0.335

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does not include T2ET16 data. See Table C.4 for a version of this table that includes T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

3.2.4 Baseline test scores

In order to estimate the effect of tutoring on test scores across baseline test scores we use the following specification:

$$Y_{isgd,t} = \alpha_0 + \sum_{i=0}^{10} \beta_i T_s \times c_i + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t}, \tag{5}$$

where c_i is the decile of the student's test score in math in T3ET15. We have 11 categories for c_i: 10 deciles and a category for those students with missing test scores. Figure 4 shows the estimates for all the βs which correspond to the treatment effect for students in a given category. Note that the effect for students with missing test scores is similar to the average treatment effect (0.06 SD). Although the treatment effect is positive for all students (except for students in the top 10% at baseline for which there is a very small, insignificant, negative effect), students in the middle of the distribution benefit more from the math tutoring (0.15 SD compared to the average effect of 0.06 SD).

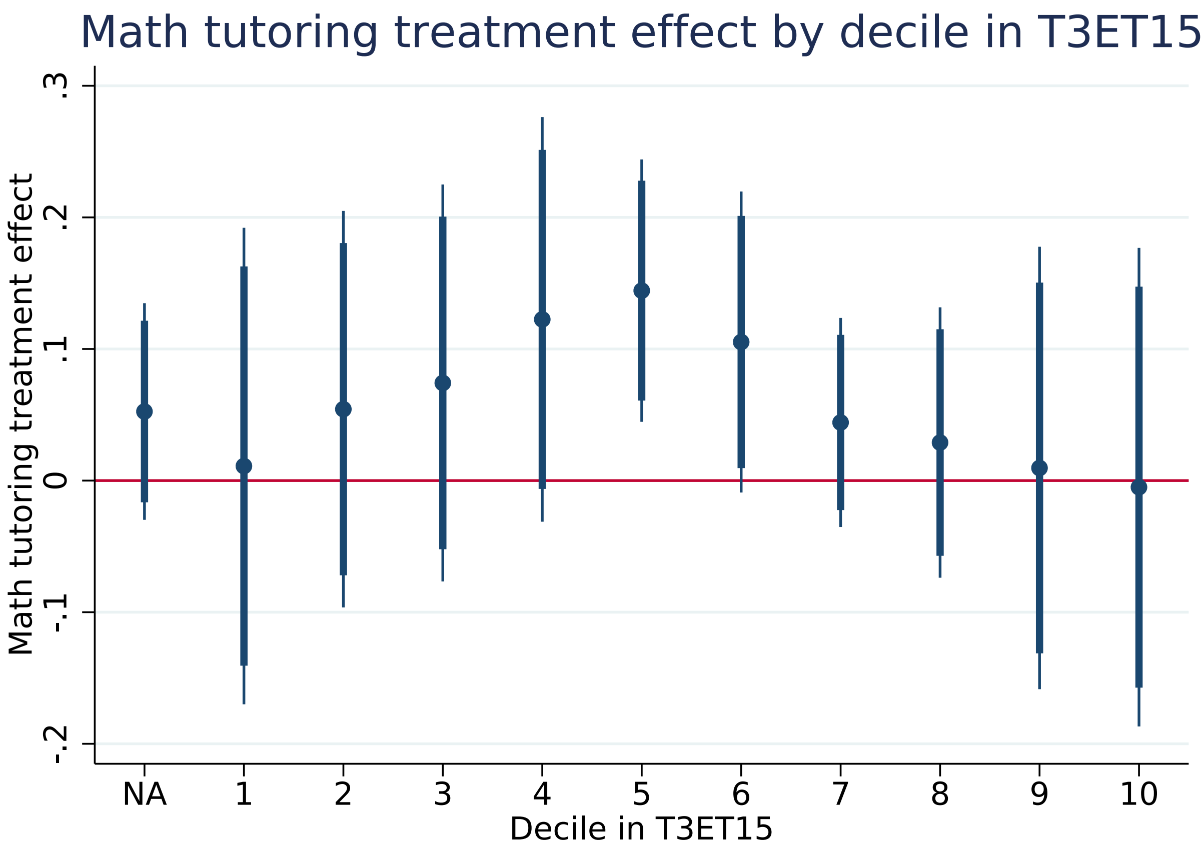


Figure 4: Treatment effect of math tutoring, relative to English tutoring, on math test scores by ability decile in T3ET15. Bars represent 90% and 95% confidence intervals (thick lines and thin lines, respectively).

3.2.5 Tutor characteristics

In order to estimate the effect of tutoring on test scores across tutor characteristics we use the following specification:

$$Y_{isgd,t} = \alpha_0 + \beta_1 T_s + \beta_2 T_s \times c_s + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t}, \tag{6}$$

where c_s is the average tutor characteristics (in school s) along which we wish to measure heterogeneity and β_2 allows us to test whether there is any differential treatment effect. Since we do not know how teachers matched students we can only measure the average characteristic of all the possible tutors a tutee might have (e.g., all the Standard 5 students for Pre-Unit tutees). Tables 13 shows that there is no differential effect by tutors' average age, gender, or baseline test score (a PCA index across all subjects).

Table 13: Heterogeneity by tutors' characteristics

	Math	English	Swahili
Panel A: Age			
Math tutoring	0.055*	-0.017	-0.042
	(0.033)	(0.035)	(0.065)
Tutors' average age	-0.096**	-0.034	-0.14*
	(0.044)	(0.046)	(0.079)
Math tutoring × Tutors' average age	0.025	0.051*	0.14*
	(0.029)	(0.029)	(0.074)
Observations	50575	48348	32805
Adjusted R^2	0.221	0.293	0.335
Panel B: Gender			
Math tutoring	0.052	-0.012	0.043
	(0.034)	(0.036)	(0.048)
Proportion of male tutors	0.029	0.015	0.049
	(0.027)	(0.024)	(0.035)
Math tutoring × Proportion of male tutors	-0.023	-0.029	-0.079*
	(0.030)	(0.027)	(0.040)
Observations	50575	48348	32805
Adjusted R^2	0.220	0.292	0.335
Panel C: Test score			
Math tutoring	0.075**	0.0031	0.035
	(0.033)	(0.037)	(0.051)
Tutors' median test score in T3ET15	0.019	-0.015	-0.022
	(0.033)	(0.037)	(0.051)
Math tutoring × Tutors' median test score in T3ET15	-0.028	0.015	0.065
	(0.038)	(0.040)	(0.057)
Observations	40892	39030	26576
Adjusted R^2	0.230	0.302	0.345

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does not include T2ET16 data. See Table C.5 for a version of this table that includes T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

3.2.6 School characteristics

In order to estimate the effect of tutoring on test scores across tutors characteristics we use the following specification:

$$Y_{isgd,t} = \alpha_0 + \beta_1 T_s + \beta_2 T_s \times c_s + \alpha_1 Y_{isgd,t=0} + \gamma_g + \gamma_t + \gamma_d + \alpha_2 X_i + \alpha_3 X_s + \varepsilon_{isd,t}, \tag{7}$$

where c_s is a school characteristic along which we wish to measure heterogeneity and β_2 allows us to test whether there is any differential treatment effect. Tables 14 shows that there is no differential effect by the tutors' pupil-teacher ratio (PTR), tutee-tutor ratio (TTR) or school size (number of enrolled students).

Table 14: Heterogeneity by school characteristics

	Math	English	Swahili
Panel A: PTR			
Math tutoring	0.054	-0.016	0.029
	(0.033)	(0.035)	(0.046)
PTR	-0.026	-0.00077	0.0024
	(0.032)	(0.031)	(0.038)
Math tutoring × PTR	0.025	-0.00054	0.045
	(0.036)	(0.034)	(0.043)
Observations	50935	48705	33119
Adjusted R^2	0.220	0.292	0.336
Panel B: TTR			
Math tutoring	0.053	-0.016	0.011
	(0.032)	(0.032)	(0.046)
TTR	-0.032	-0.11***	-0.11**
	(0.028)	(0.036)	(0.043)
Math tutoring × TTR	0.033	0.12***	0.13***
	(0.028)	(0.036)	(0.046)
Observations	50914	48684	33098
Adjusted R^2	0.220	0.295	0.337
Panel C: School size			
Math tutoring	0.054	-0.016	0.027
	(0.033)	(0.035)	(0.046)
Enrollment	-0.026	-0.0012	0.0012
	(0.034)	(0.032)	(0.038)
Math tutoring × Enrollment	0.026	0.0028	0.045
	(0.038)	(0.035)	(0.043)
Observations	50935	48705	33119
Adjusted R^2	0.220	0.292	0.336

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does not include T2ET16 data. See Table C.6 for a version of this table that includes T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

4 Conclusions

There is an increasing wealth of evidence showing that teaching appropriate to the student's learning level can improve learning outcomes in low-income countries.. However, teachers often lack the time (or incentives) to give each child personalized instruction tailored to their needs and providing schools with extra teachers to do so is expensive. Cross-age tutoring, where older students tutor younger students, is an inexpensive alternative to providing personalized instruction to younger students in that it substitutes a trained instructor (the teacher) with an untrained one (the older student) at the cost of the older student's time.

We present results from a large RCT (over 180 schools, 20,000 tutees, and 20,000 tutors) in Kenya, in which schools are randomly selected to implement a cross-age tutoring program in either English or math. We show that cross-age tutoring in math, relative to tutoring in English, has a small positive effect (0.06 SD, p-value of 0.073) on math test scores. These results do not hold true for English tutoring, however: relative to math tutoring, it has no positive effect on English test scores (we can rule out an effect of 0.077 SD with 95% confidence). We show that there is considerable heterogeneity by the students' baseline learning levels. Specifically, the effect is largest for students in the middle of the ability distribution (0.144 SD, p-value of 0.005), while the point estimates are almost zero for students with either very low or very high baseline learning levels. This is consistent with: a) tutors not being able to help students who are advanced learners and need an instructor with a high level of expertise to guide them through more advanced concepts; and b) tutors not being able to help tutees lagging behind grade level competencies who may need more specialized instruction to catch up. Surprisingly, we find no heterogeneity by other tutee characteristics (e.g., age and gender), by school characteristics, or by tutor characteristics.

Assuming that tutoring does not harm tutors in non-academic dimensions[11], our results indicate that schools are not be using their resources efficiently. Cross-age tutoring represents a "free-lunch": a net improvement for all the parties involved.

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the*

[11]It may actually yield benefits in terms of non-academic traits such as social skills.

American Statistical Association, *103*(484), 1481-1495. Retrieved from http://dx.doi.org/10.1198/016214508000000841 doi: 10.1198/016214508000000841

Bold, T., Kimenyi, M. S., & Sandefur, J. (2013). Public and private provision of education in kenya. *Journal of African Economies*, *22*(suppl 2), ii39-ii56. Retrieved from http://jae.oxfordjournals.org/content/22/suppl_2/ii39.abstract doi: 10.1093/jae/ejt014

Bruhn, M., & McKenzie, D. (2009, October). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, *1*(4), 200-232. Retrieved from http://www.aeaweb.org/articles?id=10.1257/app.1.4.200 doi: 10.1257/app.1.4.200

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014, September). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.104.9.2633 doi: 10.1257/aer.104.9.2633

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237-248. Retrieved from http://aer.sagepub.com/content/19/2/237.abstract doi: 10.3102/00028312019002237

Evans, D. K., & Popova, A. (2016). What really works to improve learning in developing countries? an analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*(2), 242-270. Retrieved from http://wbro.oxfordjournals.org/content/31/2/242.abstract doi: 10.1093/wbro/lkw004

Filmer, D., Hasan, A., & Pritchett, L. (2006). A millennium learning goal: Measuring real progress in education. *Center for Global Development Working Paper*(97).

Glewwe, P., & Muralidharan, K. (2016). Chapter 10 - improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In S. M. Eric A. Hanushek & L. Woessmann (Eds.), (Vol. 5, p. 653 - 743). Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780444634597000105 doi: http://dx.doi.org/10.1016/B978-0-444-63459-7.00010-5

Global monitoring report. (2007). The World Bank. Retrieved from http://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-6975-3 doi: 10.1596/978-0-8213-6975-3

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, *90*(5), 1184-1208. Retrieved from http://www.aeaweb.org/articles.php?doi=10.1257/aer.90.5.1184 doi: 10.1257/aer.90.5.1184

Hanushek, E. A., & Wößmann, L. (2007). The role of education quality for economic growth. *World Bank Policy Research Working Paper*(4122).

Jones, S., Schipper, Y., Ruto, S., & Rajani, R. (2014). Can your child read and count? measuring learning outcomes in East Africa. *Journal of African Economies*. Retrieved from http://jae.oxfordjournals.org/content/early/2014/06/12/jae.eju009.abstract doi: 10.1093/jae/eju009

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*(3), 1071-1102. Retrieved from http://restud.oxfordjournals.org/content/76/3/1071.abstract doi: 10.1111/j.1467-937X.2009.00536.x

Mbiti, I. M. (2016, August). The need for accountability in education in developing countries. *Journal of Economic Perspectives*, *30*(3), 109-32. Retrieved from http://www.aeaweb.org/articles?id=10.1257/jep.30.3.109 doi: 10.1257/jep.30.3.109

Munley, V. G., Garvey, E., & McConnell, M. J. (2010). The effectiveness of peer tutoring on student achievement at the university level. *The American Economic Review*, 277–282.

Quintilianus, M., & Halm, K. (1869). *Institutio oratoria* (No. v. 2). Teubner. Retrieved from https://books.google.com/books?id=jaI9AAAAcAAJ

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681-704. Retrieved from http://qje.oxfordjournals.org/content/116/2/681.abstract doi: 10.1162/00335530151144131

Shenderovich, Y., Thurston, A., & Miller, S. (2015). Cross-age tutoring in kindergarten and elementary school settings: A systematic review and meta-analysis. *International Journal of Educational Research*.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., ... Jimenez, E. (2016). The impact of education programmes on learning and school participation in low-and middle-income countries.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement*. *The Economic Journal*, *113*(485), F3–F33.

United Nations. (2015). *The millennium development goals report 2015*. United Nations Publications.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, *85*(1), 9–23.

A Extra tables

B Extra figures

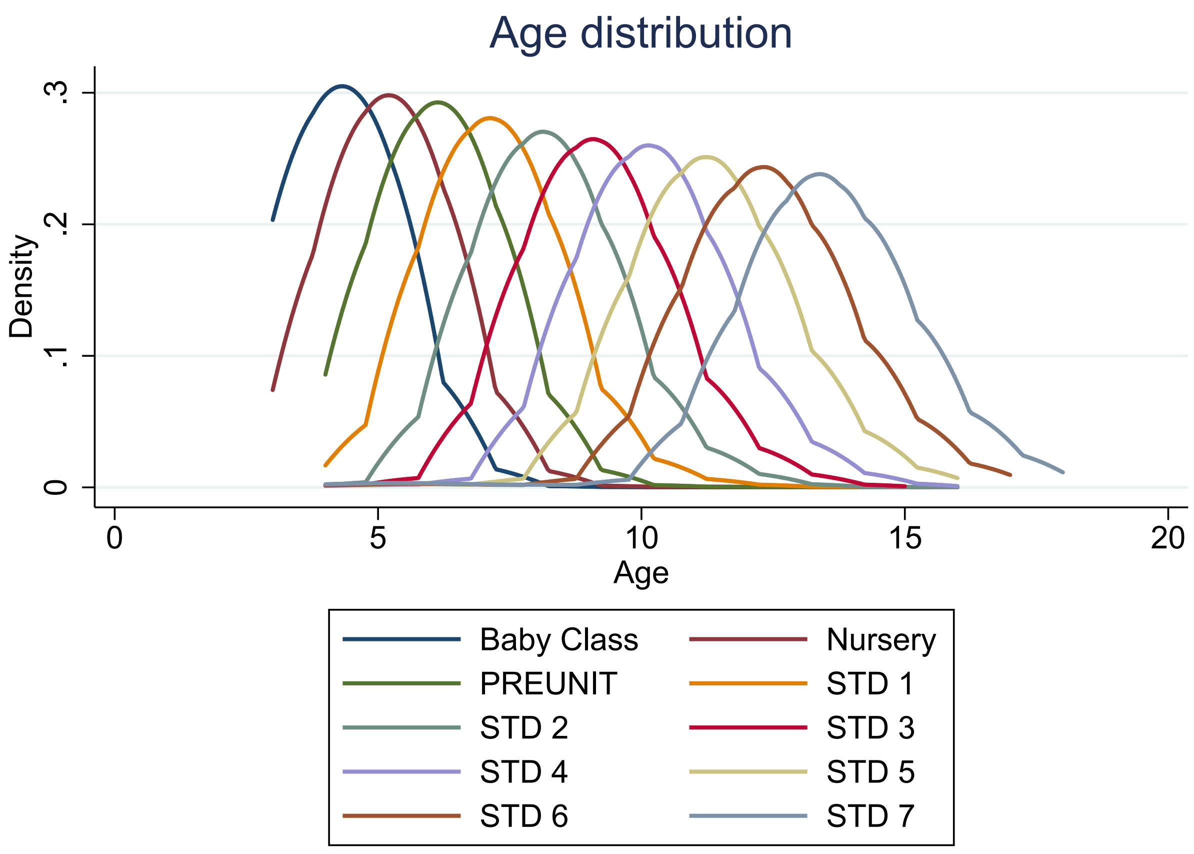


Figure B.1: Age distribution across grades

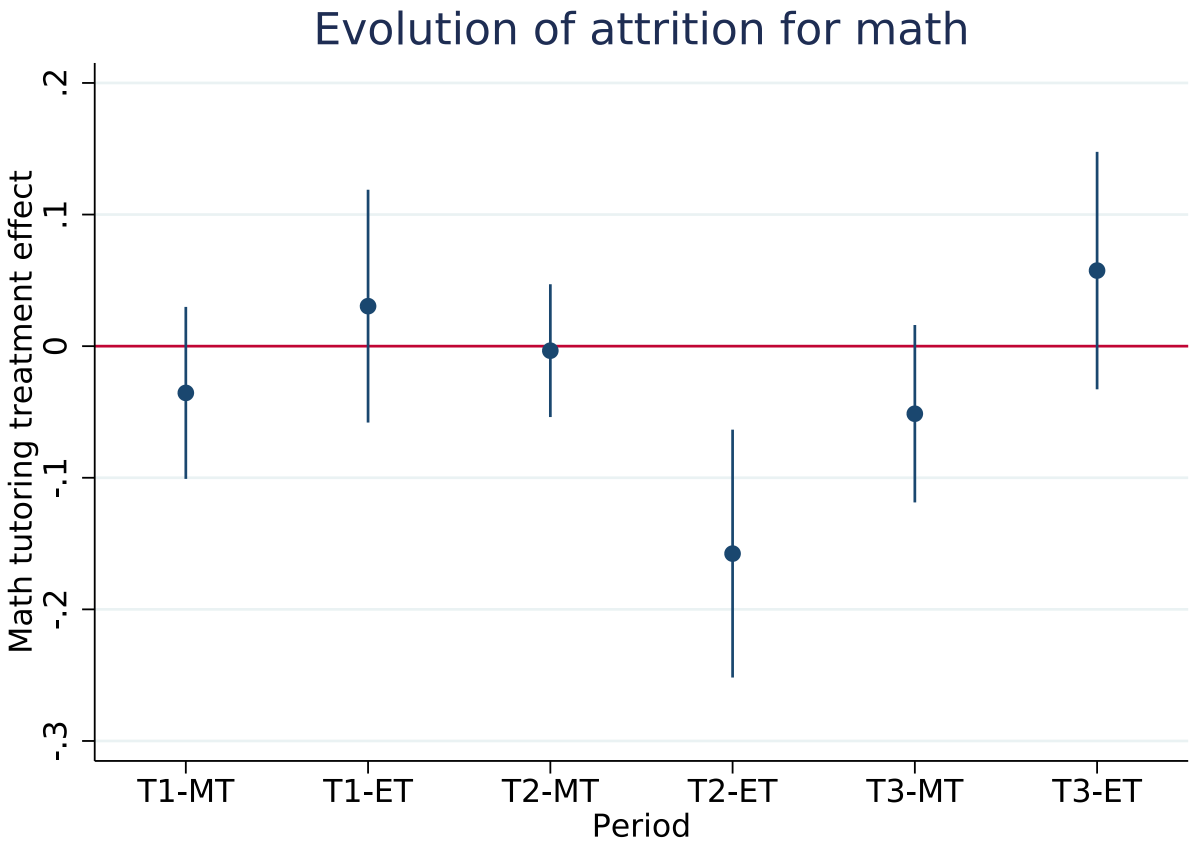


Figure B.2: Difference in testing rates across English and math tutoring in each period. Bars show 95% confidence intervals.

C With T2ET16

Table C.1: Differential attrition between treatment and control students

	Tested
Math tutoring	-0.0097
	(0.019)
Mean English	0.39
N. of obs.	100914
Number of schools	187

Clustered standard errors, by school, in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table shows the differential attrition between students in math tutoring schools compared to students in English tutoring schools. The estimation data set does include T2ET16 data.

Table C.2: Effect on tutees' test scores

	percentscore_Z
Math tutoring	0.022
	(0.050)
N. of obs.	26364
Number of schools	185

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does include T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

Table C.3: Effect on tutees' test scores

	percentscore_Z
Math tutoring	-0.014
	(0.036)
N. of obs.	52560
Number of schools	187

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does include T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

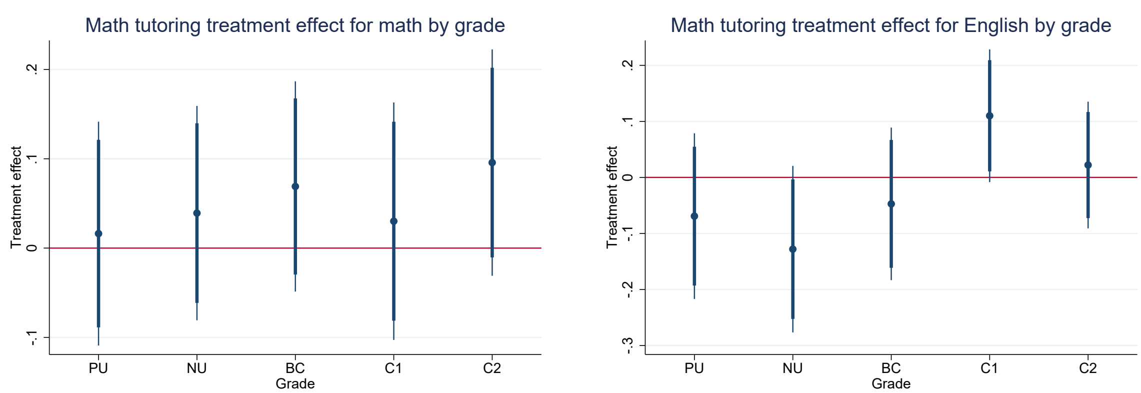


Figure C.1: Treatment effect of math tutoring, relative to English tutoring, in math (left panel) and English (right panel) test scores by grade. Bars represent 90% and 95% confidence intervals (thick lines and thin lines, respectively).

Table C.4: Heterogeneity by tutees' characteristics

	Math	English	Swahili
Panel A: Age			
Math tutoring	0.053	-0.012	-0.010
	(0.034)	(0.034)	(0.052)
Date of birth	-0.047**	0.033	0.074**
	(0.021)	(0.023)	(0.036)
Math tutoring × Date of birth	-0.028	-0.052**	-0.054
	(0.022)	(0.025)	(0.044)
Observations	57259	56364	38201
Adjusted R^2	0.222	0.293	0.336
Panel B: Gender			
Math tutoring	0.050	-0.013	0.016
	(0.033)	(0.034)	(0.046)
Male	0.0047	-0.029**	-0.020
	(0.012)	(0.013)	(0.014)
Math tutoring × Male	-0.015	-0.0029	-0.011
	(0.014)	(0.015)	(0.016)
Observations	57391	56491	38260
Adjusted R^2	0.220	0.293	0.336
Panel C: Bridge enrollment age			
Math tutoring	0.053	-0.013	0.0030
	(0.034)	(0.034)	(0.048)
AgeFirstBridge	0.017	-0.063***	-0.055*
	(0.020)	(0.020)	(0.029)
Math tutoring × AgeFirstBridge	0.034	0.036	0.037
	(0.021)	(0.023)	(0.033)
Observations	57259	56364	38201
Adjusted R^2	0.222	0.293	0.336

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does include T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

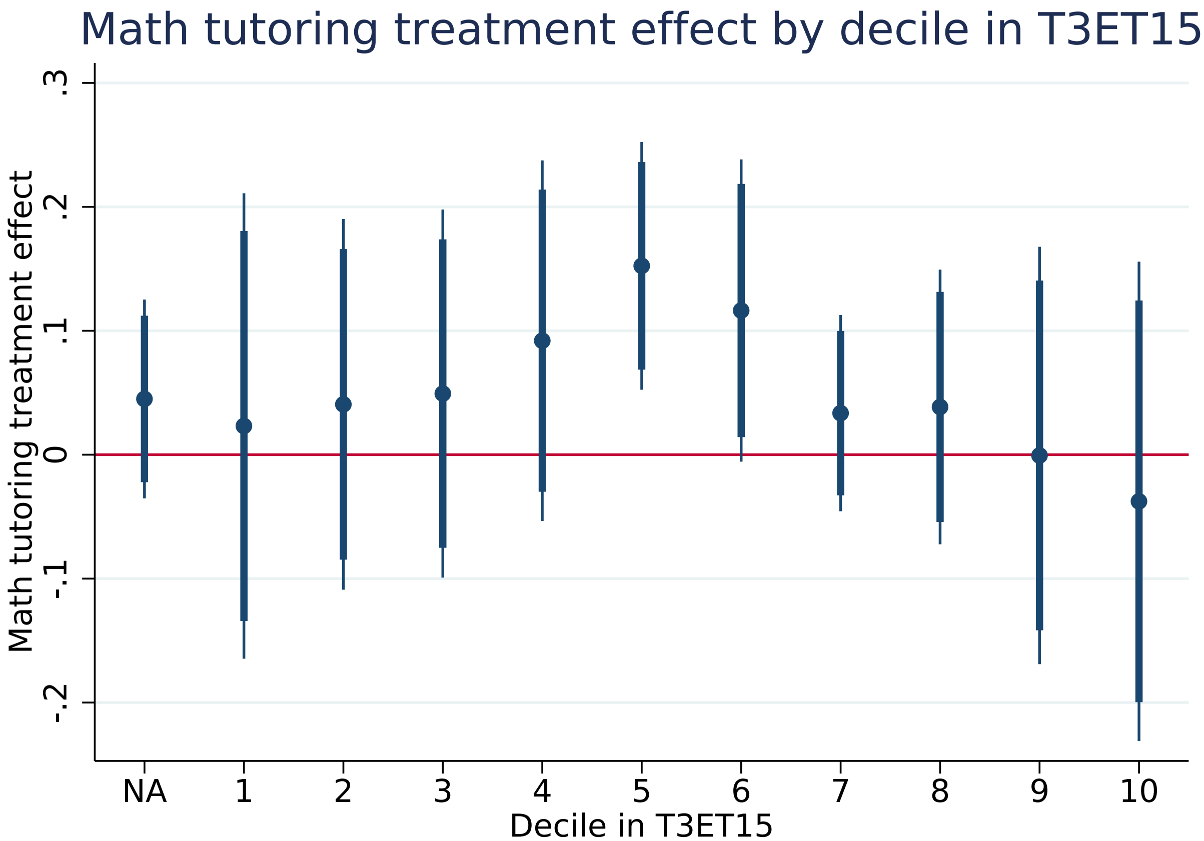


Figure C.2: Treatment effect of math tutoring, relative to English tutoring, on math test scores by ability decile in T3ET15. Bars represent 90% and 95% confidence intervals (thick lines and thin lines, respectively).

Table C.5: Heterogeneity by tutors' characteristics

	Math	English	Swahili
Panel A: Age			
Math tutoring	0.052	-0.014	-0.037
	(0.033)	(0.034)	(0.067)
Tutors' average age	-0.094**	-0.026	-0.098
	(0.045)	(0.044)	(0.081)
Math tutoring × Tutors' average age	0.016	0.049*	0.097
	(0.029)	(0.027)	(0.075)
Observations	57003	56101	37913
Adjusted R^2	0.221	0.292	0.335
Panel B: Gender			
Math tutoring	0.048	-0.0093	0.024
	(0.035)	(0.035)	(0.048)
Proportion of male tutors	0.029	0.017	0.048
	(0.024)	(0.023)	(0.033)
Math tutoring × Proportion of male tutors	-0.017	-0.029	-0.086**
	(0.029)	(0.026)	(0.040)
Observations	57003	56101	37913
Adjusted R^2	0.221	0.292	0.336
Panel C: Test score			
Math tutoring	0.072**	0.0100	0.0099
	(0.033)	(0.036)	(0.052)
Tutors' median test score in T3ET15	0.0100	-0.019	-0.053
	(0.032)	(0.036)	(0.050)
Math tutoring × Tutors' median test score in T3ET15	-0.020	0.018	0.086
	(0.038)	(0.039)	(0.056)
Observations	46347	45514	30843
Adjusted R^2	0.232	0.303	0.347

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does include T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***

Table C.6: Heterogeneity by school characteristics

	Math	English	Swahili
Panel A: PTR			
Math tutoring	0.052	-0.012	0.0086
	(0.034)	(0.034)	(0.047)
PTR	-0.029	-0.011	0.0100
	(0.032)	(0.031)	(0.038)
Math tutoring × PTR	0.026	0.0085	0.040
	(0.037)	(0.033)	(0.044)
Observations	57391	56491	38260
Adjusted R^2	0.221	0.292	0.337
Panel B: TTR			
Math tutoring	0.050	-0.014	-0.011
	(0.033)	(0.031)	(0.046)
TTR	-0.036	-0.11***	-0.12***
	(0.027)	(0.033)	(0.039)
Math tutoring × TTR	0.039	0.12***	0.13***
	(0.027)	(0.033)	(0.041)
Observations	57364	56464	38233
Adjusted R^2	0.221	0.295	0.338
Panel C: School size			
Math tutoring	0.052	-0.012	0.0065
	(0.033)	(0.034)	(0.047)
Enrollment	-0.028	-0.011	0.0063
	(0.035)	(0.032)	(0.038)
Math tutoring × Enrollment	0.027	0.010	0.041
	(0.039)	(0.034)	(0.044)
Observations	57391	56491	38260
Adjusted R^2	0.221	0.292	0.337

The independent variable is the standardized test score (mean 0 and standard deviation of 1 in English tutoring schools). Student and school controls include student's gender and age, monthly academy fees, dummies for teachers' wage categories and the pupil-teacher ratio in T1DG16. A flexible third-order polynomial is used to control for lagged test scores. The estimation data set does include T2ET16 data. Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***