NBER WORKING PAPER SERIES

DECENTRALIZING EDUCATION RESOURCES: SCHOOL GRANTS IN SENEGAL

Pedro Carneiro
Oswald Koussihouèdé
Nathalie Lahire
Costas Meghir
Corina Mommaerts

Working Paper 21063
http://www.nber.org/papers/w21063

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2015

Financial support was provided by the Education Program Development Fund (EPDF) of the World Bank. Pedro Carneiro thanks the financial support from the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (grant reference RES-589-28-0001), the support of the European Research Council through ERC-2009-StG-240910 and ERC-2009-AdG-249612. Costas Meghir thanks the ISPS and the Cowles foundation at Yale for financial assistance. Corina Mommaerts thanks the NSF Graduate Research Fellowship for support. All errors and views are ours alone and do not necessarily reflect the views and opinions of the World Bank, the funding bodies or those commenting on the paper. This RCT was registered in the American Economic Association Registry for randomized control trials under Trial number 671. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

Decentralizing Education Resources: School Grants in Senegal
Pedro Carneiro, Oswald Koussihouèdé, Nathalie Lahire, Costas Meghir, and Corina Mommaerts
NBER Working Paper No. 21063
April 2015
JEL No. H52,I22,I25,O15

ABSTRACT

The impact of school resources on the quality of education in developing countries may depend crucially on whether resources are targeted efficiently. In this paper we use a randomized experiment to analyze the impact of a school grants program in Senegal, which decentralized a portion of the country's education budget. We find large positive effects on test scores at younger grades that persist at least two years. We show that these effects are concentrated among schools that focused funds on human resources improvements rather than school materials, suggesting that teachers and principals may be a central determinant of school quality.

Pedro Carneiro
Department of Economics
University College
Gower Street
London WC1E 6BT
United Kingdom
p.carneiro@ucl.ac.uk

Oswald Koussihouèdé
University Gaston Berger
St. Louis, Senegal
okoussihouede@gmail.com

Nathalie Lahire
World Bank
Washington, DC
nlahire@worldbank.org

Costas Meghir
Department of Economics
Yale University
37 Hillhouse Avenue
New Haven, CT 06511
and IZA
and also NBER
c.meghir@yale.edu

Corina Mommaerts
Department of Economics
Yale University
37 Hillhouse Av
New Haven, CT 06511
corina.mommaerts@yale.edu

1 Introduction

In the last 50 years, primary school enrollment has increased dramatically in the developing world. Even in the poorest areas of Sub-Saharan Africa, gross enrollment rates in primary school are approaching 80 percent (e.g., Glewwe and Kremer (2006)). There is, however, widespread evidence that the quality of education in developing countries remains very low. As a result, increases in school enrollment may not translate into corresponding increases in productivity and wellbeing. This is consistent with recent evidence suggesting that education quality, not quantity, matters most for growth (e.g., Hanushek and Woessmann (2010), Glewwe et al. (2013)).

We address the following question: is it possible to improve the quality of poor schools by providing them with cash transfers? The appeal of this idea lies in its simplicity. The assumption behind it is that local decision makers, such as principals and community leaders, are likely to have a deeper understanding of the needs of their schools than central education authorities, and are therefore in the best position to put these resources to their most efficient use.

We study a school grant program in Senegal, which was developed to decentralize at least a small part of the country's education budget. Through this program, every elementary school in Senegal could apply for funds for a specific school project that seeks to improve the quality of learning and teaching, with the best proposals being selected through a competitive process. The maximum amount a school could receive for a project amounted to USD$3,190, which corresponded to 7 percent of the total annual school budget of a typical school (inclusive of teacher salaries).

We find large and statistically significant effects on test scores one year after the start of the intervention, for children who benefited from school grants when they were in second grade - especially for girls. The effects are larger for schools in the South of the country, where projects tended to focus on training human resources (teaching and management), compared to the North, where priority was placed on the acquisition of school material (e.g., textbooks/manuals). We do not observe similar program impacts for children in other grades. The point estimates are very similar in the second follow up for the same children, pointing to persistent effects.

Since we examine the impact of the intervention across different tests and different groups of students, for inferential purposes we implement a step-down procedure proposed by Romano and Wolf (2005) that controls the probability of falsely rejecting at least one true null hypothesis,

and improves upon more conservative prior methods for multiple hypothesis testing such as the Bonferroni procedure. We show that our main conclusions survive and are unlikely to be due to false rejections.

The evidence on the effect of school resources on primary school student achievement in developing countries is at best mixed (see Glewwe and Kremer (2006), Glewwe et al. (2013), and Murnane and Ganimian (2014) for reviews). While some pedagogical resources, such as textbooks and flipcharts, only have positive effects for high-achieving students (see Glewwe et al. (2009), Glewwe et al. (2004)), other resources such as computer-assisted instruction increased test scores by up to one-half of a standard deviation in India (Banerjee et al. (2007)). If local decision-makers can target resources better than a central authority, however, school grants (and other ways of decentralizing funding) could help boost the effect of school resources by targeting funds toward efficient uses of resources (see Galiani and Perez-Truglia (2013) for a review).

In theory, decentralizing decision making about school resources is a good idea, but in practice, this idea is often dismissed. The main problem is that principals in public schools (or other local decision makers) may not be incentivized to use resources optimally to serve students. Recent work on secondary schools in Argentina and primary schools in the Gambia find positive effects of decentralization (Galiani et al. (2008), Blimpo et al. (2014)). Meanwhile, cross-country comparisons show negative effects of decentralization for developing countries (Hanushek et al. (2013)). Our results indicate that decentralized distribution of resources through school grants can have positive effects on student achievement, and we present suggestive evidence that factors such as teacher quality may have enhanced the impacts.

The paper proceeds as follows. In Section 2 we describe the school grants program in Senegal and the evaluation design. In Section 3 we describe our data and Section 4 describes our empirical approach. In Section 5 we present our main results and examine potential mediating factors through which the impact of the program may have operated. Section 6 concludes.

2 Description of the Program and Evaluation

Primary schooling in Senegal consists of six years of education and is funded through a mix of government, foreign aid, and household resources.[1] Almost all classroom instruction is conducted in French, while the language spoken by students at home is predominantly not French (only 11 percent of the household interviews were conducted in French). Gross enrollment rates in primary schools increased dramatically over the ten years prior to our study, from 67 percent in 2000 to 92 percent in 2009. Despite this large increase in enrollment, in 2009 only 60 percent of students completed primary school. In an effort to increase the quality of primary education, Senegal's Ministry of Education initiated this school grants program (see Appendix A for more historical details).

2.1 School Grants in Senegal

For the past several years, Senegal has used school grants (projets d'école) as a tool to fund improvements in education quality, based on the premise that school-level actors are in the best position to identify a school's unique deficiencies and the most workable solutions to address them. Beginning in 2009, the emphasis of these grants shifted from strengthening the physical environment toward pedagogic issues. At that point the government also sought technical and financial support from the World Bank to rigorously evaluate the program.

The main goal of the program was to improve school quality, as measured by student learning outcomes, specifically by improving pedagogical resources in the school. Instead of providing general funding for all schools, funds were targeted towards problems identified by the school as major obstacles to quality, and identified by a government evaluation committee (Inspection Départementale de l'Education Nationale, IDEN) as being eligible for funding based on district-level and system-wide priorities. Problems were identified at the local level, in the hope that decentralized decision-making would allow more efficient and effective use of funds.

Generally, the program worked as follows. The Ministry of Education issued a call for proposals, based on the available grant funding, priority areas, and eligible activities (and sometimes eligible regions). Schools that decided to apply for funding completed a grant application for a school

[1] Fees collected from parents represent around ten percent of school funding in 2006 (PASEC (2007)) and are a non-trivial financial burden on families: around one-fifth of students who dropped out in the first year of primary school did so because of limited financial resources of their parents (World Bank (2013)).

project (called the projet d'école) addressing a particular pedagogical issue faced by the school. Another important component of the program was its role in promoting strong community participation in schools. As a result, grants were prepared by a committee of parents, teachers, and local officials. For schools that received a grant, the grant totaled around 1,500,000 CFA Francs (approximately USD$3,190), which represented a roughly 7 percent increase in expenditures per student in a typical school (inclusive of teacher salaries, which comprise over 90 percent of the budget).[2] In Section 2.2 we describe the process through which grants were approved and allocated.

2.2 Evaluation Design

In the initial stage of this study, all Senegalese schools were eligible to respond to the call for proposals. The IDEN evaluation committee first ranked the applications and discarded low quality and ineligible applications. The remaining ones, referred to as "approved applications" were grouped into two categories. The first consisted of very good proposals which were eligible for financing. The second consisted of strong proposals with potential, but which needed revision. These were sent back to schools with comments from the IDEN evaluation committee, then re-submitted. Figure 1 provides a graphical representation of this process.

This process resulted in the selection of 633 projects to fund, whose locations are shown in Figure 2.[3] For the purposes of the evaluation, these 633 projects were randomly allocated to three funding cohorts. 211 schools were selected randomly to receive funding in the first cohort (June 2009), at the end of the school year. This funding could only be executed at the beginning of the following school year (October/November). Of the remaining schools, 211 were to receive funding in June 2010, and another 211 were to receive funding in June 2011. In practice, the disbursement of the second round of grants did not occur until the first trimester of 2011. This means that between mid-2009 and mid-2011, two groups of schools can be compared. The schools in the first cohort received school grants during this period, while the schools in the second and third cohorts did not and therefore can be used as a comparison group for the schools in the first cohort. The school year runs from October/November through June, allowing us to compare the first cohort to

[2]These numbers are based on collected self-reports from principals and teachers in our sample.

[3]Of these projects, 96 percent included a component to improve French outcomes, 70 percent had a component to improve math outcomes, and 52 percent had a component to improve science outcomes. 82 percent of the projects aimed to build capacity, 63 percent aimed to increase teaching time, and 45 percent aimed to reduce repetition and drop-out. The intended beneficiaries of these projects, in addition to students, were the teachers and principal in 84 percent of projects, and the management committee in 29 percent of projects.

Figure 1: Evaluation Design

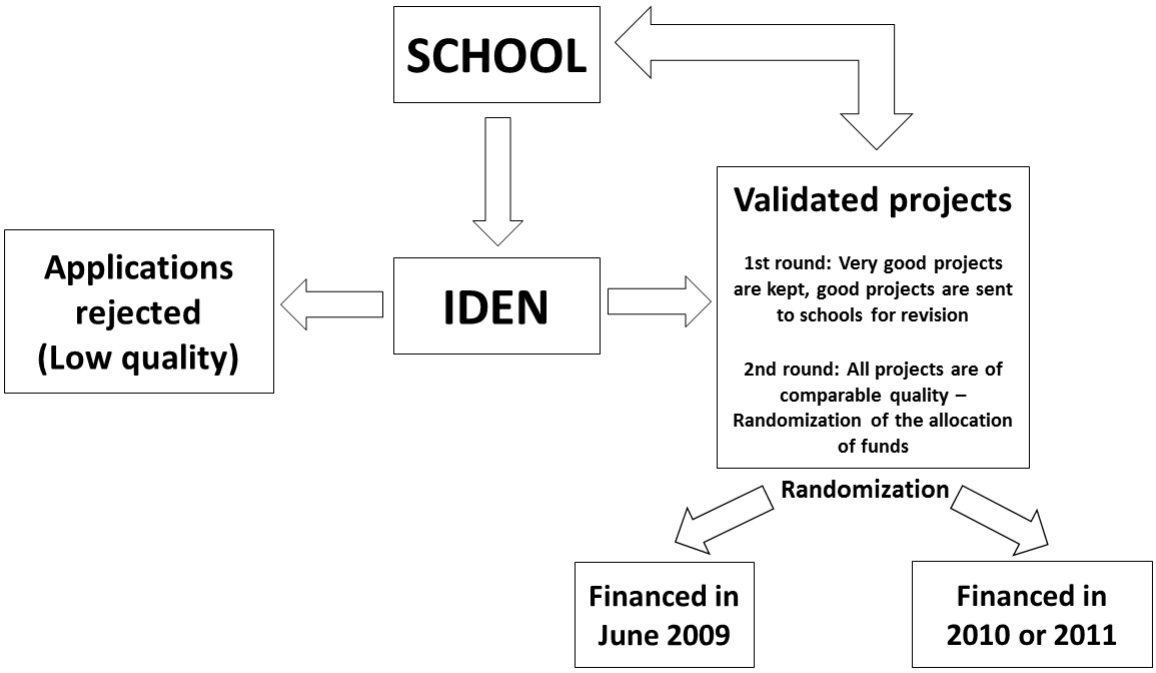


Figure 2: Location of Schools in Sample

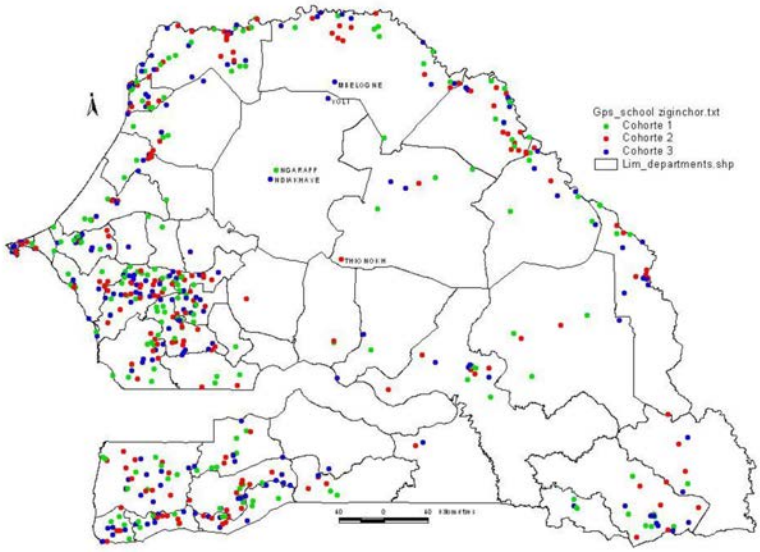
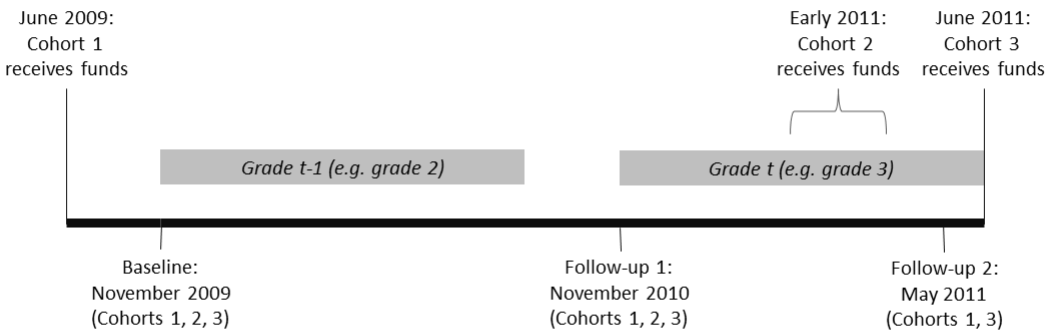


Figure 3: Evaluation Timeline



both the second and third cohorts for the 2009-2010 school year and the first cohort to the third cohort for the 2010-2011 school year (see Figure 3).

The randomization among eligible schools is critical for our study: it ensures that the three successive cohorts are statistically comparable, which in turn ensures unbiased estimates of the effect of the program. In this process it is crucial that the control group contains only schools that were judged as eligible but were not selected to receive funding by the randomization process until a later date.

3 Data and Balance

In order to gather data for this study, three waves of surveys were administered to students and their families, teachers, and principals in these schools. A baseline survey was conducted at the start of the 2009-2010 academic year (in November), right as the first round of grants were able to be executed. Subsequent surveys took place in November 2010 at the beginning of the 2010-2011 academic year (first follow-up), and in May 2011 at the end of the 2010-2011 academic year (second follow-up).

At baseline, we administered written assessments in mathematics and French to a random sample of 6 children in each of grades 2 and 4, and an oral reading assessment (similar to Early Grade Reading Assessment, or EGRA) to a random sample of 3 of those 6 children in grades 2 and 4. Importantly, the same tests were administered across all waves. In addition, we randomly selected 2 of the 3 children in each grade who took all three assessments, and conducted a household survey that included demographic and financial information on all household members. Finally, we collected classroom and school level information by surveying the school principals and the teachers of the students in our sample.

In the first follow-up, we surveyed and tested the same children again (at the start of 3rd and 5th grade, respectively) and their households, teachers and principals. Schools who received grants in the first cohort answered a set of questions on the use of the extra funds. To examine the possibility that funds were disproportionately channelled to students preparing to enter secondary school, we also administered written assessments in mathematics and French to a random sample of children who were in 6th grade at follow-up, and also surveyed their teachers.

In the second follow-up, we re-surveyed and tested the same children who were tested at baseline and first follow-up, and we also tested a random sample of 2nd, 4th, and 6th grade students in French and mathematics. In addition, in the second follow-up we administered the Peabody Picture Vocabulary Test (PPVT) to children and their mothers. We did not collect general school and classroom information in the second follow-up.

Of the 633 schools, split randomly into three cohorts of 211 schools each, we sampled 525. We were able to contact 478 schools at baseline (among which 447 were successfully surveyed), 528 at first follow-up[4] (among which 517 were successfully surveyed), and 340 at second follow-up (among which 325 were successfully surveyed and tested).[5] The schools that were not included at baseline were out of bounds either due to inclement weather or rebel activity in the South. While this may have impacted the representativeness of the baseline sample, it did not affect the balance as accessibility was not correlated with treatment status, as we will report later. Due to budgetary

[4]We contacted more schools in the first follow-up than we originally sampled because the enumerators accidentally went to an extra treatment and two extra control schools that we had not originally planned on sampling.

[5]See Appendix Table A15 for the corresponding number of student-level observations and attrition. In Appendix Tables A16 and A17 we show the difference in baseline characteristics between treatment and control schools, for students who did not leave the sample between baseline and first follow-up or second follow-up, respectively. The sample is similarly balanced as our main sample (see below).

constraints, in the second follow-up we dropped Cohort 2 schools, and ended up with a sample of 352 schools, of which 325 schools were successfully surveyed and tested. Since cohorts were randomly allocated, this did not introduce bias.

The first column of Panel A of Table 1 shows means and standard deviations of baseline test scores for females in grade 2 of control schools. The third column shows the same results for boys (analogous results for grade 4 are in Appendix Table 1, and the full distribution of scores is in Appendix B). The French, mathematics, and oral tests were appropriately targeted to student grade level, as mean scores (calculated as the proportion of correct responses on the exam) were around 30-40 percent. The same tests were administered at first follow-up, so these scores allowed room for noticeable improvement. The fourth row corresponds to an index of the three tests (which is the first principal component of these three tests, standardized to have unit variance).

Panel B shows household characteristics of the students. On average these students live less than a kilometer from the school and miss one day of school per month. Their households spend a fair amount of their income on education expenses as compared to household food consumption, and over half of the parents claim to be involved in school activities. Only 10 percent of the household interviews were conducted in French.

Panel C reports school characteristics. The average school in our sample is not small: it has 347 students and 10 teachers, half of whom hold a baccalaureate degree and half of whom participated in training in the five years preceding the intervention. The schools are varied in their resources: 56 percent have electricity, and 23 percent have a library. Three-quarters of principals have a baccalaureate degree.

The second and fourth columns in panels A and B show the differences in test scores and school characteristics between treatment and control at baseline, for both males and females. The second column of panel C shows differences in household characteristics between treatment and control. All but one difference (parental involvement in school) are insignificant. Thus treatment and control are very well balanced. It is noteworthy that the precision of the difference in test scores is very high, which bodes very well for our ability to detect even small effects of the program.[6]

[6]With the exception of the index score, we chose not to standardize the mathematics, French, and oral scores. The tests were designed to appropriately measure the types of skills taught in the first years of elementary school, and looking at the proportion of right answers in this test is a natural way to assess student knowledge in these subjects, and its progress over time. Furthermore, these scores are specific to Senegal, so standardization would not be useful for international comparisons. Even within sample, we show below that the distribution of scores is highly

Table 1: Baseline Descriptive Statistics and Balance (grade 2)

	Females		Males	
	Control	Treat-Control	Control	Treat-Control
Panel A: Test Scores				
Percent Correct: French	0.413	-0.001	0.435	-0.010
	(0.219)	(0.018)	(0.225)	(0.019)
Percent Correct: Math	0.351	-0.006	0.382	-0.001
	(0.223)	(0.018)	(0.234)	(0.020)
Percent Correct: Oral	0.212	0.011	0.235	0.009
	(0.165)	(0.016)	(0.174)	(0.018)
Index Score (standardized)	-0.085	0.012	0.081	-0.012
	(0.952)	(0.096)	(0.991)	(0.104)
Panel B: Household Characteristics				
Days of school missed last week	0.209	0.057	0.128	0.070
	(1.085)	(0.102)	(0.580)	(0.081)
Student works after school	0.010	0.016	0.010	-0.003
	(0.100)	(0.014)	(0.099)	(0.009)
Household size	9.365	0.223	9.301	-0.241
	(4.048)	(0.423)	(4.065)	(0.432)
Number of children in household	5.211	0.183	5.350	-0.277
	(2.568)	(0.269)	(2.665)	(0.288)
Head has any education	0.605	-0.059	0.586	0.037
	(0.490)	(0.050)	(0.493)	(0.051)
Percent of adult females with any education	0.366	-0.038	0.363	-0.010
	(0.399)	(0.041)	(0.419)	(0.043)
Distance to school (km)	0.697	-0.035	0.742	-0.108
	(0.821)	(0.077)	(0.987)	(0.086)
Parent involved in school	0.369	0.141***	0.401	0.003
	(0.483)	(0.050)	(0.491)	(0.051)
Expenditure on household food (1,000s CFA)	21.806	1.385	21.828	1.273
	(16.099)	(1.524)	(14.830)	(1.533)
Expenditure on uniform (1,000s CFA)	2.442	0.378	2.423	-0.131
	(1.395)	(0.461)	(0.966)	(0.385)
Expenditure on tuition (1,000s CFA)	1.114	-0.017	1.107	0.003
	(1.091)	(0.113)	(1.279)	(0.117)
Expenditure on supplies (1,000s CFA)	3.460	0.151	4.281	-0.906*
	(2.475)	(0.267)	(7.614)	(0.497)
Student has tutor	0.148	0.007	0.157	-0.026
	(0.355)	(0.038)	(0.365)	(0.036)
Home has electricity	0.478	0.038	0.471	0.016
	(0.500)	(0.052)	(0.500)	(0.052)
Home has modern toilet	0.528	-0.012	0.572	-0.026
	(0.500)	(0.053)	(0.496)	(0.051)
Land owned (hectares)	2.234	0.262	2.508	0.618
	(3.104)	(0.451)	(3.789)	(0.813)
Interview conducted in French	0.114	-0.030	0.121	-0.049
	(0.318)	(0.029)	(0.327)	(0.031)

Continued on next page

Table 1, continued: Baseline Descriptive Statistics and Balance (grade 2)

	All Students	
	Control	Treat-Control
Panel C: School and Teacher Characteristics		
Distance to nearest city (km)	18.378	-0.067
	(25.007)	(2.176)
Locality population (100,000s)	1.383	0.035
	(4.400)	(0.459)
Locality has health center	0.709	0.026
	(0.454)	(0.043)
School located in South	0.185	-0.007
	(0.388)	(0.037)
School has Electricity	0.566	0.013
	(0.496)	(0.048)
Number of Teachers	9.678	0.439
	(4.966)	(0.511)
Number of Pupils	341.113	28.473
	(252.387)	(25.604)
School has library	0.206	0.079*
	(0.405)	(0.042)
Number of computers	1.281	-0.011
	(4.386)	(0.397)
Number of manuals in classroom	59.897	3.173
	(45.183)	(4.581)
Percent teachers female	0.316	0.010
	(0.235)	(0.023)
Average teacher age	33.118	-0.132
	(4.242)	(0.389)
Percent of teachers with Baccalaureate	0.413	-0.019
	(0.227)	(0.023)
Average teacher experience	6.558	0.083
	(3.693)	(0.350)
Percent teachers with training in past 5 years	0.474	0.096**
	(0.499)	(0.049)
Percent of principals with Baccalaureate	0.735	-0.051
	(0.441)	(0.044)

Columns 2 and 4 report the difference in means between treatment and control. Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As we explained above, some schools were inaccessible at baseline, and thus were only added to the survey in the first follow up (although they participated in the randomization, and the treatment schools in this group were funded as planned). The exclusion from baseline was unrelated to treatment status, which explains why nevertheless baseline schools are balanced. In Appendix Table A2 we present descriptive statistics for all schools including those added at the first follow up. As we expect, when we compare the characteristics of treatment and control schools which we did not expect to change as a result of the experiment there is no significant difference, other than possibly in distance from school. However this is just one significant difference among many differences; jointly there are no differences and this one is very small in magnitude. Hence, whether we look at schools surveyed at baseline or at the first followup, there is no evidence of imbalances between treatment and control with respect to their baseline characteristics.

Another concern is that these 633 schools may be fundamentally different from other primary schools in Senegal as a result of the grant selection process (e.g., these schools were better organized to put together a good grant application). Thus, they may not constitute a random set of schools in Senegal and the results of this study may not generalize. In Appendix Table A3, we show characteristics of a nationally representative sample of Senegalese households using data collected in 2006 by PASEC (Programme on the Analysis of Education Systems), a survey aimed at assessing educational attainment in primary school, and variables that correspond to those in our data. Schools in our sample have fewer students and are more likely to have electricity than the average school in Senegal, but are similar on other measures, including the literacy rates, the number of teachers and their education, and whether the school has a library. At least in terms of these variables, our sample does not look drastically different from the average Senegalese primary school.

4 Empirical Approach and Inference

We use a regression approach to estimate the impacts of the program. Specifically, the impacts are the estimated β_t^k coefficients[7] from the following regression:

$$Y_{ist}^k = \alpha_t^k + \beta_t^k G_s + X_{is}\lambda_t + \varepsilon_{ist}^k \quad (1)$$

non-normal, so a one standard deviation in test scores does not have the usual meaning. Nevertheless, in Appendix C we replicate our main results using standardized test scores.

[7] Appendix Table A4 reports results using standardized test scores.

where Y_{ist}^k is the proportion of correct answers in test k, for student i in school s at follow up t (1 or 2), G_s is a treatment indicator, X_{is} are conditioning variables measured at baseline, and ε_{ist}^k is an error term. Conditioning variables include household size, number of children, whether the head has any education, distance to school, a wealth index[8], the interview language, and the baseline scores of all tests. Since household interviews were conducted for only a random subsample of students, two-thirds of our sample has missing household characteristics (at random). In order to keep these observations, we assign zeros to conditioning variables if they are missing and include dummies for observations with missing conditioning variables.[9]

Since we are testing multiple hypotheses at once we compute levels of significance for each coefficient using the step-down approach of Romano and Wolf (2005). In this way we control for the family-wise error rate (FWE). The FWE is defined as the probability of incorrectly identifying at least one coefficient as significant, which becomes more likely as the number of hypothesis tests increases. The Romano-Wolf approach improves upon more conservative classical methods such as the Bonferroni correction by applying a "step-down" algorithm that takes advantage of the dependence structure of individual tests. Our approach is to control for a FWE of 5 and 10 percent and mark each coefficient that is significant at each of these rates with ** and * respectively. However, testing too many hypotheses at once may reduce power to detect anything significant. We thus test multiple hypotheses in related groups rather than for all effects reported in the paper.

We also present bootstrap-based 95% confidence intervals, clustered at the school level. They may exclude zero, pointing to a coefficient that is significant at a 5% level, while the multiple-testing adjusted p-value may be above 5% or even above 10%. This is perfectly consistent because the p-value is adjusted for multiple testing, while inference based on the 95% CI only corresponds to a 5% level of significance if no other hypothesis is tested.

[8]The wealth index is standardized to have unit variance and is defined as the first principal component of the following variables: the home has electricity, the home has plumbing, the home has a radio, the home has a television, the home has a telephone, the home has a computer, the home has a refrigerator, the home has gas, the home has an iron, the home has a bicycle, the home has an automobile, the home has a bed, the home has a modern toilet, the number of chickens, the number of sheep, the number of cows, the number of horses, the number of donkeys, the amount of land, savings, debt, food expenditure, child expenditure, other expenditure, wall material, ground material, and roof material.

[9]Results without conditioning variables are presented in Appendix C and they are almost identical, but of course less precise.

5 Results

5.1 Impacts on Test Scores for Boys and Girls

In Table 2 we present estimates of the difference between third grade test scores in treatment and control schools, measured at first follow-up (beginning of third grade) in Panel A, and at second follow-up (end of third grade) in Panel B.

As explained above, at first follow-up we have measurements of student performance in written tests in French and mathematics, as well as an oral test that covers sound, letter and word recognition, and reading comprehension, but (for cost reasons) was only administered to a third of the students who take written tests. For each of these three tests we compute the proportion of correct answers given by each student. In addition, we use the first principal component as a summary index of these three tests, which is standardized to have mean zero and standard deviation 1. For the second follow-up, we also have scores for the Peabody Picture Vocabulary Test, which is standardized to have mean zero and standard deviation 1 (within sample).

Before proceeding to a more detailed analysis of the impacts, we first note that the effects of the program for third grade students on the entire set of 42 main outcomes we consider[10] are jointly significant with a p-value of 0.02. This conclusion is based on a χ^2 type test with critical values derived using the bootstrap. Thus the intervention had an overall significant effect on third grade outcomes.

We start with the results for children who were in third grade at the first follow up, one year after the intervention started. Panel A of Table 2 has 4 columns, one for each of the three tests, and one for the aggregate index of all tests. There are two rows. The first presents the estimates for female students and the second presents estimates for males. Each cell displays estimates of β_t^k, the corresponding 95% bootstrap confidence interval, and significance stars based on the Romano and Wolf (2005) procedure.

There are large impacts of school grants on third grade test scores of girls. This is true across all tests. Impacts range from an increase of over 3 (math) to almost 5 (oral) percentage points in the proportion of correct answers in each test. These are large effects in light of the means (and

[10]The outcomes jointly tested include the 3 tests for boys and girls in both followups plus the PPVT in the second follow-up. We consider a national sample (14 hypotheses), and a sample split between the north and south of the country (28 hypotheses), for a total of 42 hypotheses.

Table 2: Program Impacts on Grade 3 Test Scores

Panel A: Beginning of Third Grade (First Follow-Up)

	French	Math	Oral	Index[a]
Female	0.037	0.031	0.047*	0.217**
	[0.005, 0.068]	[0.004, 0.057]	[0.015, 0.080]	[0.071, 0.358]
Male	0.022	0.024	0.011	0.041
	[-0.009, 0.055]	[-0.004, 0.050]	[-0.024, 0.044]	[-0.100, 0.188]
Observations	2720	2718	1385	1350

Panel B: End of Third Grade (Second Follow-Up)

	French	Math	Oral	Index[a]	PPVT[a]
Female	0.043	0.019	0.054	0.246	0.096
	[0.006, 0.082]	[-0.016, 0.053]	[0.007, 0.103]	[0.057, 0.447]	[-0.155, 0.358]
Male	0.026	0.016	0.024	0.078	0.207
	[-0.008, 0.063]	[-0.019, 0.051]	[-0.018, 0.066]	[-0.099, 0.251]	[-0.016, 0.410]
Observations	1732	1721	853	826	566

[a] The index and PPVT columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ correspond to Romano Wolf (2005) p-values from a joint test (see text for included hypotheses). Conditioning variables: Household size, number of children, education of head, distance to school, wealth index, interview language, baseline scores, missing dummies. In the first follow-up, overall means of French, math, and oral tests are 0.540, 0.538, and 0.351 for females and 0.546, 0.564, and 0.376 for males, respectively. In the second follow-up, overall means of French, math, oral, and (standardized) PPVT tests are 0.678, 0.678, 0.463, and -0.030 for females and 0.685, 0.696, 0.474, and 0.025 for males, respectively.

standard deviations) of these test scores: 0.54 (0.24) for both written French and written math, and 0.35 (0.22) for the oral test. If we view all of these tests as noisy measurements of one underlying human capital factor we may improve precision by using the first principal component computed using factor analysis. When we look at this aggregate index of the three tests, the school grant increases third grade school performance of females by 0.22 of a standard deviation.

All of these effects are individually significant (i.e. the 95% confidence interval does not include zero) for girls, but none are for boys, where the effects are much smaller and the index points to no overall improvement. To evaluate the statistical significance of these effects as a group, as explained above, we first carry out a joint χ^2 test for all effects reported in Table 2 (excluding the index). Jointly these effects are significant at the 2% level. To pinpoint which effects are responsible for the overall rejection of the null hypothesis, we implement the Romano and Wolf (2005) procedure and find that the oral test for girls is significant at the 10% level but none of the other effects are significant.[11] However, the effect on the aggregate index for girls, which effectively filters out measurement error, is still significant at the 5% level once we control for multiple testing.[12] Taken together, these results imply substantial, sizable and significant effects on the performance for girls but did not change the overall performance of boys.

Panel B of Table 2 documents program impacts measured using test scores at the end of third grade for the same students (the second follow-up). In addition to the three tests and the index described above, column 5 shows the impact on PPVT scores, standardized to have mean zero and standard deviation 1 in the sample.

Impacts of the program for boys are again positive but much smaller than for girls, below 2.6 percentage points for all tests considered (about 0.08 of a standard deviation for the index). These are not significant by any criterion.

For females, impact sizes are similar to those reported in Panel A of Table 2, if not larger, with the exception of math. This indicates that program impacts persisted two years after the grant was disbursed to schools. The effects we report for girls are all individually significant, except for the PPVT and mathematics in the second follow up. However, we note that, based on the step-down

[11] Here we are including the French, mathematics, and oral outcomes for boys and girls for the first follow up, overall and by north and south of the country (18 joint hypotheses).

[12] Here the joint test includes the index for the first follow up for males and females, overall and by north and south of the country (6 joint hypotheses).

p-values, the effects are not significant in the second followup (albeit the sample used is smaller, since we could not use cohort 2 schools).[13] Our interpretation of these results is that the program seems to be improving the performance of girls who, at the start of the program, were in second grade (and tested at the beginning and the end of third grade).

It is interesting that a relatively small grant is able to improve children's learning outcomes to this extent, especially for girls. By contrast, in Glewwe and Kremer's (2006) survey of the recent literature on the effectiveness of improvements in school resources on students' learning in developing countries, there are several interventions that show no significant impact. In developed countries, there are even fewer examples of successful school resource interventions (Hanushek (2006)).

It is possible that the intervention improved outcomes because it provided cash in a decentralized way to local decision makers, who could then put these funds to an efficient use. Nevertheless, there is abundant evidence of leakages in other similar grant programs across the world (Reinikka and Svensson (2004), Bruns et al. (2011)). If the extent of local capture of these funds is also substantial in Senegal then the results in this paper are even more remarkable because they would have been produced with minimal resources.

Whether the program has different effects across the distribution is an important question relating to targeting. In Figure 4 we show parameter estimates together with their 95% confidence intervals from a quantile regression of the relevant test scores for grade three in the first follow up, on the treatment indicator and including the usual controls, clustered by school. For girls the effects of the grant are spread over most of the distribution, except at the very top. For boys these results confirm no effect along the entire distribution.

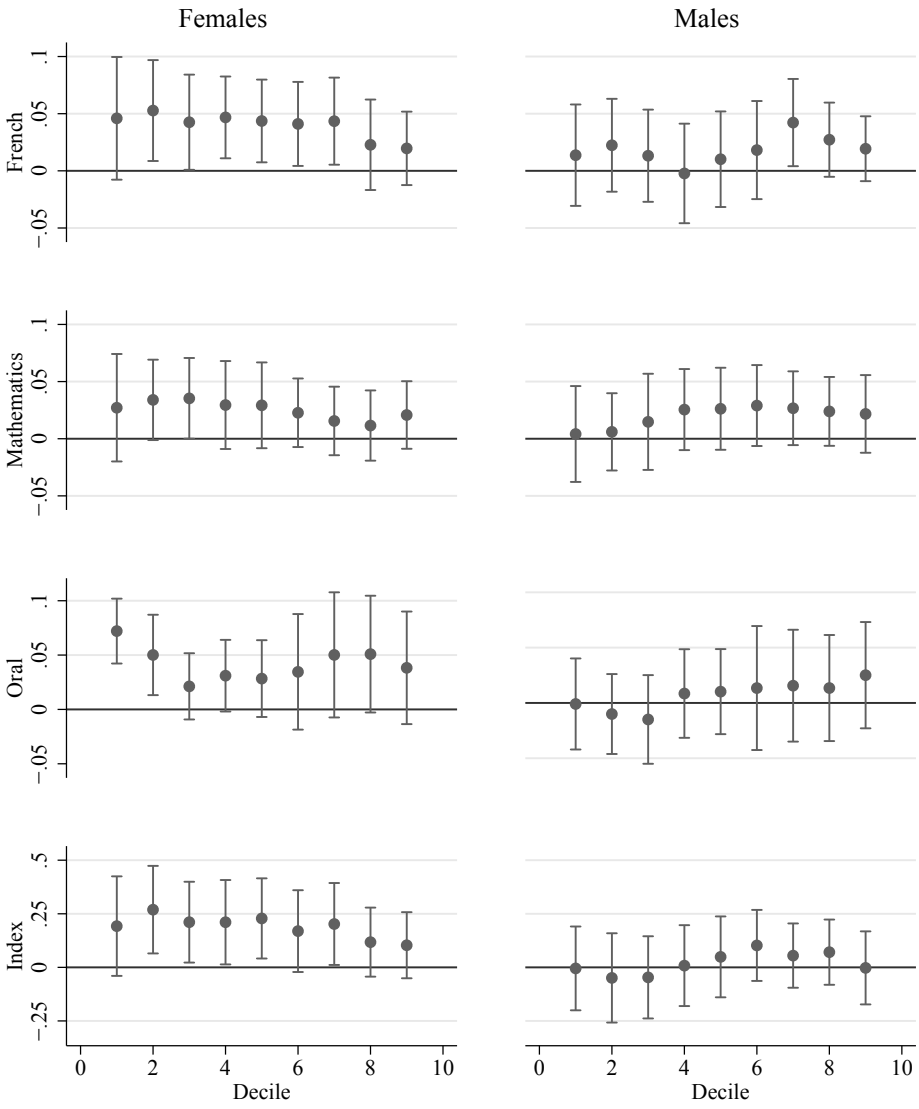
A similar picture emerges for the second follow up in Figure 5, although the results are unfortunately less precise because of the smaller sample size. Finally, estimates of the grant on PPVT scores occur mainly in the middle of the distribution, for boys. No other effect is observed for boys on the remaining test scores.

We also estimated the main effects for children who were at grade 5 in the two followups in Table 3. The impacts are numerically close to zero and statistically insignificant by any criterion.[14]

[13] Notice that all the individual confidence intervals increase relatively to those in Panel A.

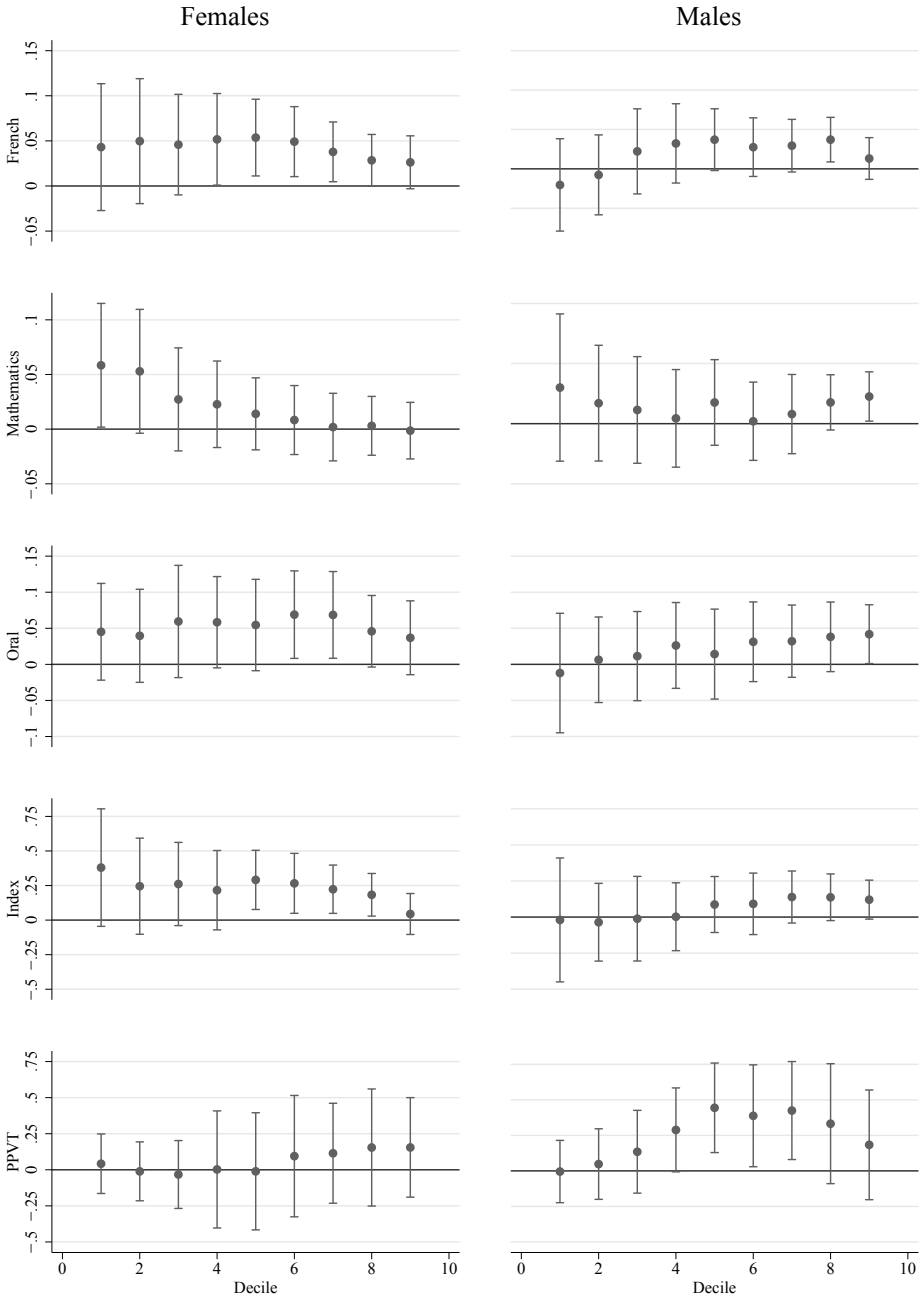
[14] Appendix Table A6 resports these effects without controls, as well as the impacts of school grants on test scores of children in other grades for which we collected test scores but no other data.

Figure 4: Distributional Impacts on Test Scores at the Beginning of Third Grade (First Follow-up)



Note: Point estimates from a quantile regression at each decile with 95% confidence intervals. Index coefficients are standardized.

Figure 5: Distributional Impacts on Test Scores at the End of Third Grade (Second Follow-up)



Note: Point estimates from a quantile regression at each decile with 95% confidence intervals. Index and PPVT plot coefficients are standardized

It is remarkable that there are only (individually) statistically significant program impacts (either positive or negative) on grade 3 tests. The standard errors of the estimates are similar across grades, but the point estimates are quite small across all tests, grades, and survey waves, with the exception of grade 3.[15]

Table 3: Program Impacts on Grade 5 Test Scores

Panel A: Beginning of Fifth Grade (First Follow-Up)

	French	Math	Oral	Index[a]
Female	0.009	0.013	-0.005	0.009
	[-0.014, 0.033]	[-0.017, 0.041]	[-0.034, 0.025]	[-0.104, 0.131]
Male	0.014	0.008	0.020	0.043
	[-0.009, 0.035]	[-0.021, 0.034]	[-0.011, 0.051]	[-0.084, 0.171]
Observations	2648	2643	1347	1329

Panel B: End of Fifth Grade (Second Follow-Up)

	French	Math	Oral	Index[a]	PPVT[a]
Female	0.008	-0.003	-0.006	-0.022	-0.096
	[-0.019, 0.033]	[-0.029, 0.023]	[-0.046, 0.030]	[-0.163, 0.108]	[-0.339, 0.155]
Male	-0.001	-0.013	0.019	0.046	-0.029
	[-0.029, 0.027]	[-0.044, 0.013]	[-0.019, 0.055]	[-0.113, 0.212]	[-0.280, 0.218]
Observations	1606	1606	833	794	556

[a] The index and PPVT columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. Conditioning variables: Household size, number of children, education of head, distance to school, wealth index, interview language, baseline scores, missing dummies. In the first follow-up, overall means of French, math, and oral tests are 0.470, 0.417, and 0.628 for females and 0.490, 0.459, and 0.659 for males, respectively. In the second follow-up, overall means of French, math, oral, and (standardized) PPVT tests are 0.590, 0.543, 0.705, and -0.022 for females and 0.601, 0.580, 0.743, and 0.030 for males, respectively.

This is a very puzzling finding. It is possible that a belief that learning delays emerge early in the life of the child, and that in the early stages of school it is central to build a strong foundation for future learning, leads principals to invest in earlier grades more than in later grades. This seems to be a widespread belief among elementary school principals in Senegal.

Using data from the teachers' questionnaires at follow-up we investigate whether there were differential impacts of school grants on observable investments in 3^{rd} and 5^{th} grade students in

[15]Therefore, the lack of statistically significant results in grade 5 (but not in grade 3) does not appear to be due to a lack of power. If the point estimates for grade 5 were as large as those for grade 3 it is likely that we would be able to reject that they were statistically equal to zero. When designing our study we anticipated that with our sample we would be able to detect program impacts of between 0.2 and 0.3 standard deviations, which is in line with what we find.

Panel A of Appendix Table A7.[16] Some of the variables we can study are classroom materials (e.g., textbooks/manuals, desks, tables, etc), and teacher training. We find no differential impact of the program in any of these. When we examine other classroom characteristics or teacher behaviors, the only interesting difference to report concerns student (mis-)behavior in the classroom. While in third grade there was a positive impact of the program on student behavior as measured by the number of times a day a teacher needs to demand silence, in fifth grade there was a negative impact of the program on student behavior measured by this variable, and by the number of times a teacher has to punish a child for impolite behavior.

It is also possible that parents believe that investments in the early grades are more productive than investments in later grades. If that were true it could happen that parents decreased their home investments (if they were substitutes with school investments) more in later grades than in earlier grades in response to an increase in school resources. However, once again, there is no evidence that this took place, at least in terms of observable parental investments (see Panel B of Appendix Table A7).

5.2 Heterogenous Impacts

There are important differences in the impact of school grants on the test scores of boys and girls. Here we consider two additional characteristics by which the impact of the school grants may plausibly differ: prior ability and region (the South is much poorer and geographically distinct from the North).

For baseline ability, we convert corresponding baseline test scores into a "high" (above median) or "low" (below median) binary variable. However, as we mentioned above, several schools were missing at baseline. In Appendix Table A8 we show that missing schools at baseline are mainly in the South, and that they display worse student performance in the first follow-up than comparable non-missing schools. It is noteworthy that they are not disproportionately control or treatment schools.

For region, we distinguish schools located in the most southern regions in Senegal (Ziguinchor and Kolda) from schools in the rest of the country. We consider these regions separately because Ziguinchor and Kolda are much poorer regions (ANSD (2007)) and have been beset by problems

[16] Ideally we would want to do this using 2^{nd} and, say, 4^{th} grade students, but we do not have the follow-up data for these teachers, although we have baseline data for them.

related to rebel activity.

Since our larger estimates of program impacts were for students in 3^{rd} grade, who were first exposed to the program in 2^{nd} grade, we focus this analysis of heterogeneous impacts on them. A similar analysis performed on the test results of students in 5^{th} grade did not produce evidence of any program impacts for this set of students (see Appendix Tables A9 and A10).

Table 4: Program Impacts on Grade 3 Test Scores by Baseline Ability

Panel A: Beginning of Grade 3 (First Follow-up)

	French		Math		Oral		Index[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
Low	-0.007	0.016	-0.013	-0.000	0.049	-0.003	0.020	-0.072
	[-0.054, 0.035]	[-0.026, 0.061]	[-0.052, 0.024]	[-0.036, 0.041]	[0.003, 0.092]	[-0.042, 0.039]	[-0.158, 0.223]	[-0.287, 0.144]
High	0.056	-0.001	0.049	0.010	0.010	-0.001	0.243	0.032
	[0.007, 0.102]	[-0.047, 0.044]	[0.009, 0.084]	[-0.025, 0.046]	[-0.036, 0.057]	[-0.048, 0.043]	[0.033, 0.434]	[-0.149, 0.201]
p-value (diff)	0.036	0.574	0.020	0.683	0.216	0.940	0.098	0.463

Panel B: End of Grade 3 (Second Follow-up)

	French		Math		Oral		Index[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
Low	0.025	0.027	0.003	-0.004	0.060	0.008	0.095	0.041
	[-0.030, 0.076]	[-0.025, 0.087]	[-0.050, 0.049]	[-0.056, 0.056]	[-0.008, 0.130]	[-0.052, 0.067]	[-0.208, 0.410]	[-0.247, 0.308]
High	0.052	0.004	0.028	0.023	0.025	0.032	0.241	0.097
	[-0.001, 0.105]	[-0.052, 0.058]	[-0.018, 0.078]	[-0.019, 0.061]	[-0.039, 0.098]	[-0.031, 0.089]	[-0.018, 0.497]	[-0.138, 0.305]
p-value (diff)	0.439	0.526	0.468	0.376	0.473	0.567	0.487	0.759

[a] The index columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or Table 2 for control variables. The P-value row shows the p-value of the difference in the two treatment effects of the column.

Table 5: Program Impacts on Grade 3 Test Scores by Region

Panel A: Beginning of Grade 3 (First Follow-up)

	French		Math		Oral		Index[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
South	0.125**	0.078	0.069	0.090*	0.102	0.048	0.491*	0.287
	[0.060, 0.194]	[0.011, 0.154]	[-0.001, 0.137]	[0.028, 0.156]	[0.045, 0.161]	[-0.031, 0.118]	[0.172, 0.775]	[-0.055, 0.581]
North	0.015	0.009	0.021	0.008	0.035	0.003	0.146	-0.016
	[-0.021, 0.051]	[-0.025, 0.045]	[-0.007, 0.051]	[-0.024, 0.040]	[-0.005, 0.071]	[-0.032, 0.038]	[-0.015, 0.302]	[-0.161, 0.152]
P-value (diff)	0.006	0.096	0.208	0.022	0.057	0.294	0.044	0.086

Panel B: End of Grade 3 (Second Follow-up)

	French		Math		Oral		Index[a]		PPVT[a]	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
South	0.088	0.067	0.066	0.095	0.143	0.054	0.547	0.282	0.192	0.213
	[0.011, 0.173]	[-0.015, 0.158]	[-0.018, 0.139]	[0.021, 0.176]	[0.043, 0.242]	[-0.042, 0.136]	[0.139, 1.049]	[-0.067, 0.642]	[-0.438, 0.802]	[-0.429, 0.853]
North	0.035	0.012	0.010	-0.008	0.031	0.014	0.170	0.018	0.098	0.199
	[-0.008, 0.078]	[-0.028, 0.052]	[-0.029, 0.045]	[-0.045, 0.028]	[-0.015, 0.081]	[-0.032, 0.063]	[-0.066, 0.395]	[-0.152, 0.206]	[-0.195, 0.385]	[-0.044, 0.422]
p-value (diff)	0.289	0.252	0.201	0.021	0.042	0.440	0.125	0.184	0.787	0.969

[a] The index and PPVT columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or Table 2 for control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ correspond to Romano Wolf (2005) p-values from a joint test (see text for included hypotheses). The P-value row shows the p-value of the difference in the two treatment effects of the column.

The regressions we run to construct Tables 4 and 5 extend equation (1) to include an interaction between the treatment variable G_s and a pre-determined variable W_{ist} (baseline ability or region):

$$Y_{ist}^k = \alpha_t^k + \beta_t^k G_s + \delta_t^k (G_s * W_{ist}) + \psi_t^k W_{ist} + \varepsilon_{ist}^k \quad (2)$$

Tables 4 and 5 have two panels, corresponding to W_{ist} ="Baseline ability" and W_{ist} ="Region", respectively. Each panel reports the treatment effect for each W_{ist} (δ_t^k) and the 95% bootstrapped confidence interval, clustered at the school level. The final row reports the p-value of the difference between the estimated treatment effects for high and low ability children.

The first panel of Table 4 shows that the impact of school grants on 3^{rd} grade French scores is especially large for girls with a high level of baseline ability in French. This is consistent with the idea that investments in skills are complementary over time and hence will be more productive for those with high levels of skill to start with. There are several education interventions that share this characteristic.

There are also several education interventions that benefit mostly girls. It is much less common to find programs that affect boys alone. Again, this could be related to the skills that girls bring to elementary schools, such as discipline, patience, and higher levels of maturity overall, which may make them better able to enjoy the benefits of additional school resources, such as a better teacher, better training manuals, a library, and so on.

We now turn to differences by North and South of the country, two very different regions. In Tables A12 and A13 of the appendix we show that the samples are balanced within each geographic region. There are dramatic differences in program impacts depending on whether the school is located in the South of the country (which are poorer and have worse school results) or in the North (Table 5, or see Appendix Table A11 for standardized coefficients). In fact, if we focus on 3^{rd} grade French scores, there are no statistically significant impacts of the program in the North of the country, whereas in the South they are very large, even among boys. For example, as a result of the program, girls in southern schools are able to increase the proportion of correct answers by 12.5 percentage points, which is over 0.5 of a standard deviation. These effects are qualitatively similar for other tests and persist through the end of the grade (second follow-up). When we examine all of the tests and correct the p-values for multiple testing, the impacts remain significant despite the high number of hypotheses tested. In addition, we still find no effects for fifth graders (Appendix

Tables A9 and A10).

The South-North differences in estimates of the impact of school grants are striking. It may be the case that the types of investments made in response to the grants varied by region and took different amounts of time to manifest themselves in test scores. In the remainder of the paper we examine whether there are differences between what school principals, teachers, and parents did in response to the availability of school grants in each of these areas, which could help shed light on the sources of regional differences in the impacts of the program on the performance of students.

5.3 Understanding Differences Between South and North

We start by examining baseline test performance differences of third grade students between schools in the South and in the North. These are shown in Table 6, panel A. Students in the southern schools perform worse on almost all tests than their counterparts in the North. This is particularly true for girls. For control schools in the first follow-up, documented in panel B, the differences between the North and the South are much larger, they hold across gender groups, and they always show larger scores in the North. A possible interpretation of this is that schools in the south had more room for improvement than schools in the North. However, this interpretation conflicts with the finding that program impacts are larger for students with high baseline test scores.

Next, we examine differences in household and project characteristics in Table 7. Panel A compares household characteristics of students in the South and in the North. Because of the missing schools at baseline, we take characteristics measured in the first follow-up among students in the control schools. A few interesting patterns emerge. Households in the south are poorer but have fewer children and better educated heads (and more prominently so for the families of female students).

Panel B of Table 7 considers the characteristics of projects being undertaken by schools with the school grant funds. This information comes from a survey conducted in treatment schools which asked principals about the project for which they got funding. We conducted two of these surveys, one at first follow-up, and one at second follow-up. We report estimates from the second follow-up survey when, presumably, data about the project is more mature and complete.

In the South, students were much more frequently named as participants in the drafting of the proposal. Although it is not clear what input students may have had, this could indicate that

Table 6: South - North Difference in Test Performance, Second-Third Grade

Panel A: Beginning of Second Grade (Baseline)

	Female	Male
Percent Correct: French	-0.015	0.035
	(0.027)	(0.031)
Percent Correct: Math	-0.068***	-0.028
	(0.024)	(0.026)
Percent Correct: Oral	-0.099***	-0.078***
	(0.017)	(0.019)
Index Score (standardized)	-0.357***	-0.226*
	(0.123)	(0.126)

Panel B: Beginning of Third Grade (First Follow-Up, Control Schools)

	Female	Male
Percent Correct: French	-0.174***	-0.132***
	(0.026)	(0.026)
Percent Correct: Math	-0.145***	-0.125***
	(0.025)	(0.024)
Percent Correct: Oral	-0.193***	-0.111***
	(0.022)	(0.027)
Index Score (standardized)	-0.894***	-0.575***
	(0.104)	(0.125)

Each coefficient reported is the difference in test score between south and north (south-north). The mean test scores in the South for French, math, and oral at baseline are 0.400, 0.292, and 0.134 for females and 0.461, 0.358, and 0.173 for males, respectively, and at first follow- up are 0.426, 0.430, and 0.215 for females and 0.457, 0.486, and 0.298 for males. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: South - North Differences in Grade 3 Household and Project Characteristics

Panel A: Household Characteristics, First Follow-up	
	Females
Household size	-1.945***
	(0.594)
Number of children in household	-0.743**
	(0.366)
Head has any education	0.118*
	(0.070)
Percent of adult females with any education	-0.033
	(0.050)
Wealth index	-0.829***
	(0.112)
Interview conducted in French	0.066
	(0.051)
Panel B: Project Characteristics, Second Follow-up	
	All Students
Months since project began	-7.564***
	(1.144)
Students helped draft application	0.253***
	(0.082)
Project included manuals	-0.095
	(0.074)
Project included computer materials	-0.092**
	(0.042)
Project included teacher training	0.162**
	(0.062)
Project included management training	0.261***
	(0.093)
Project included building courses	0.151***
	(0.046)
Project included improving general education	0.106
	(0.100)
Project included improving educational outputs	-0.015
	(0.063)
Amount spent on principal (1,000,000s CFA)	0.048***
	(0.014)
Amount spent on teachers (1,000,000s CFA)	0.039
	(0.058)
Amount spent on management (1,000,000s CFA)	0.087***
	(0.022)
Amount spent on students (1,000,000s CFA)	-0.520***
	(0.092)

	Males
	-1.408**
	(0.546)
	-0.365
	(0.385)
	0.157**
	(0.078)
	0.072
	(0.062)
	-0.778***
	(0.122)
	0.115*
	(0.059)

Each coefficient reported is the difference in test score between south and north (south-north). Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

principals were more sensitive to the needs of the students in the South. It is also significant that projects in the South started later. By the end of year 2 of the study projects in the North had been running 7.6 months longer than in the South. If results faded out quickly this could explain why we observe effects of the more recent projects than in the earlier projects but this is unlikely to be the case, given our previous results about the sustainability of program impacts (although those are not very precise). If, on the other hand, a project needed time before it started to influence children's learning (as in the case of activities that take time, such as training a teacher, or building a library), we would expect larger impacts for more mature projects, which goes against what we find in terms of the South - North comparison.

Some of the most remarkable differences relate to the components of the project. The schools in the North were more likely to have components involving the purchase of textbooks/manuals and other educational materials, while schools in the South were much more likely to have components related to training of teachers and of the management committee. This is also reflected in the amounts schools reported the project spent on principals, teachers, the management committee, and the students. Thus there are clear differences in the characteristics of projects in schools in the North and the South, as stated by the principals of these schools. Schools in the South seem to be investing more in the teaching and management abilities of their human resources, while schools in the North invest more in materials. This may well be a force behind the large differences in program impacts in these two sets of schools.

Table 8 reports the impact of the program on principals' (panel A) and teachers' (panel B) behaviors. We present separate estimates of program impacts in the South and in the North, and test whether differences in program impacts in these two areas are equal to zero (column 3).

As expected, there are no broad impacts of the school grants on aspects of school infrastructure. This was expected because, as we mentioned above, the projects had to have exclusively a pedagogical emphasis. However one aspect that can be considered infrastructure was very significantly affected by school grants both in the North and in the South: the existence of a library in the school. While the impact is twice as large in the South as in the North, we cannot reject that the two impacts are statistically equal. In addition, schools in the North that received a school grant spent more money on electricity and water for the school.

Regarding school materials and training, we see that the school grants caused an increase in

Table 8: Program Impacts on School Characteristics by Region, First Follow-Up

	South	North	South-North
Panel A: School Characteristics			
Age of youngest infrastructure	1.135	0.298	0.837
	(1.500)	(0.928)	(1.764)
Number of teachers	-0.975	0.823	-1.798
	(1.085)	(0.564)	(1.223)
Number of students	-29.100	51.321*	-80.421
	(49.317)	(29.009)	(57.216)
School has library	0.201**	0.120**	0.081
	(0.086)	(0.049)	(0.099)
Number of books in library	15.343	85.753*	-70.410
	(80.607)	(44.284)	(91.971)
Amount spent on infrastructure	40.337	53.156	-12.819
	(39.718)	(40.678)	(56.853)
Amount spent on electricity/water	-10.421	29.550*	-39.972**
	(7.867)	(15.735)	(17.592)
Amount spent on manuals	27.388**	23.019**	4.369
	(11.112)	(10.507)	(15.293)
Amount spent on tutoring	50.230*	13.512*	36.718
	(29.365)	(7.731)	(30.366)
Amount spent on teacher training	30.487**	27.856*	2.630
	(13.825)	(14.315)	(19.901)
Teacher composition changed in past year	-0.201**	-0.064	-0.138
	(0.086)	(0.042)	(0.096)
Percent teachers female	-0.031	0.012	-0.043
	(0.040)	(0.025)	(0.048)
Average teacher age	0.273	0.315	-0.041
	(0.772)	(0.432)	(0.885)
Percent of teachers with Baccalaureate	-0.043	-0.008	-0.035
	(0.049)	(0.025)	(0.056)
Average teacher experience	0.224	0.098	0.126
	(0.583)	(0.396)	(0.705)
Panel B: Third Grade Teacher Characteristics			
Minutes spent preparing lesson	3.226	1.894	1.332
	(2.941)	(2.061)	(3.591)
Number of manuals	10.475	4.990	5.486
	(7.647)	(4.980)	(9.125)
Number of measuring instruments	0.805**	-0.039	0.844**
	(0.338)	(0.208)	(0.397)
Times per day ask for silence	-5.060***	-0.774	-4.287**
	(1.847)	(0.699)	(1.975)
Times per day punish a student	0.263	-0.242	0.505
	(0.792)	(0.268)	(0.837)

Column 3 reports the difference between the impact of the program in the south and the north. Clustered standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Amounts in CFA.

books in the library in the North and an increase in the amount spent on manuals in both regions. In contrast, schools in the South spent substantially more in tutoring while both sets of schools increased spending on teacher training. All this is very much consistent with the way principals described the grant projects, as reported in Table 7. While the point estimates reveal differences in direction in the North and South, it is difficult to be conclusive since none of the impacts are significantly different between the two (except expenditure on electricity and water).

It is also interesting that there was an increase in the number of students in the North (perhaps attracted by more school materials) which is not matched by an equally large increase in the number of teachers, and which could lead to a dilution of treatment effects. In the South both these quantities go down, but not significantly. Nevertheless, it is not easy to explain how these changes in class size came about.

Finally, school grants decreased teacher turnover, particularly in the South. Given that teachers are likely to be the most important input in the school production function, the fact that in the South the program significantly affected the amount of training they got and how likely they were to remain in the school from one year to the next, is consistent with the finding of strong program impacts on student performance in this region of the country.

One final and important thing to report is that, as a result of the grant, the behavior of students seems to have improved considerably in the South, but not in the North: treatment affected how often teachers have to ask for silence during the day in southern schools. This mimics what we found before when we compared 3^{rd} and 5^{th} grade teacher reports (see Appendix Table A7). Student behavior improved among 3^{rd} graders but not among 5^{th} graders, which is exactly what happened in terms of test results. It is difficult to say whether test results improved because teachers were able to improve student behavior, or whether teachers who are good at teaching reading and math are also able to improve behaviors in their students.

We also examined the impact of the program on household behaviors in the South and in the North, which is shown in Appendix Table A14. It could happen that households in the North reacted differently to the availability of school grants than those in the South, say, by investing less (substitution) in their children (which would then help explain the pattern of test results). However, this did not happen. There are no noteworthy impacts of the program on household behaviors, and they do not seem to vary with the region of the country where households are located.

The resulting picture from this section is mixed. There are several differences between the south and the north: households are poorer yet more educated households in the south, and projects in the south tended to focus on training and less on information technology. However, when we look at the impact of the grant on how schools use the money, there are no obviously significant differences between the north and south. Nevertheless, the improvement in behavior in the south is remarkable and one can expect that fewer classroom disruptions - perhaps due to teacher training - can help learning.

6 Conclusions

There is substantial debate about the importance of resources in schools for student performance. More often than not, increases in school resources are not associated with increases in student performance. One reason may be that central education authorities lack an understanding of the needs of schools. Principals, on the other hand, could have better information and could target resources more efficiently. The danger is that incentives to improve student performance may vary across school principals and there may be several sources of local pressures for alternative uses of these funds.

This paper studies the impact of a school grant program on student performance and on potential mechanisms that could underlie the change in school performance induced by such a program. We find impacts of school grants on student learning, especially on girls with high ability levels at baseline. Notably, these impacts persist over the two years of our evaluation. However, these impacts occur only in third grade, as opposed to later grades, and they are stronger in the South of the country. These results suggest that resources distributed in a decentralized manner can have positive impacts on students.

While it is difficult to explain the grade differential in program impacts, one conjecture is that principals focus on earlier grades because they see there the foundations for future learning. We can say a bit more, however, about the North-South difference in program impacts, based on how we see principals spending their resources. While schools in the North emphasized manuals and other education materials, schools in the South emphasized human resources, namely through the training of teachers and school administrators. Our results suggest that the latter type of investments, although perhaps less visible to the local community (and therefore less preferred by

say, local politicians, or even local school authorities), is likely to be more effective than the former type of investments. This result is also consistent with the idea that the main determinant of school quality is teachers, not equipment, as suggested by the most recent literature on this topic (e.g., Hanushek and Rivkin (2006)).

Online Appendix: Not for Publication

A. Primary Education in Senegal: School Grants Program

Senegal is investing a large portion of public expenditures in education (6 percent of GDP in 2009). About 90 percent of the budget is geared towards teacher salaries. Yet, despite these resources, results have been slow to materialize. While access to primary education has increased (gross enrollment rates reached 92 percent in 2009), primary completion rates remain low (57 percent in 2009).

Senegal's education decentralization policy dates back to 1972. A decree in 2002 established school management committees (Comite de Gestion des Ecoles), which are composed of representatives of the village chief, the commune, the Parents' Association, the local sports and cultural association, school directors, teachers, and others. The CGEs' central role is to manage and supervise the implementation of school grants. Other education management aspects, such as teacher salaries, continue to be steered at the central level.

B. Test Score Distributions

In this appendix we document the way the distribution of test scores changed over time and as a result of the experiment. The first panel of Figure A1 shows the densities of scores in the French written test taken by girls in 2nd grade at baseline, first follow-up, and second follow-up. The dashed lines in the figures correspond to the control schools while the solid lines are the treatment schools. The second panel of Figure A1 shows the corresponding figure for boys. As expected given earlier results, the thinnest, lightest lines show that the distributions of test scores are balanced for treatment and control at baseline (beginning of second grade) for children of both genders.

The lines with a medium level of thickness correspond to the densities of scores at the beginning of third grade at first follow-up. Again, the dashed lines correspond to control schools, and the solid lines correspond to treatment schools. Although the densities of test scores are balanced at baseline, they are different at first follow-up, especially for girls, indicating that the program has an impact on this group of children. Notice also that students in the first follow-up have a much higher (and statistically significant) proportion of correct answers than at baseline (recall that they take the same test in both occasions). Given that the baseline takes place at the beginning of second grade, and the first follow-up at the beginning of third grade, this is expected if schools provide knowledge about the test material. Notice also that program impacts are smaller for boys.

The thickest lines in Figure A1 correspond to the densities of scores for the second follow-up, conducted at the end of third grade. We use the same test as in the previous two waves. Like before, the evolution of test scores shows (statistically significant) learning during third grade. Program impacts remain strong for girls towards the end of the third grade, roughly two years after the funds were disbursed to the treatment schools.

Figures A2 and A3 show that this is also true if we examine math and oral tests.

Figure A1: Distribution of Second/Third Grade French Scores, by Gender

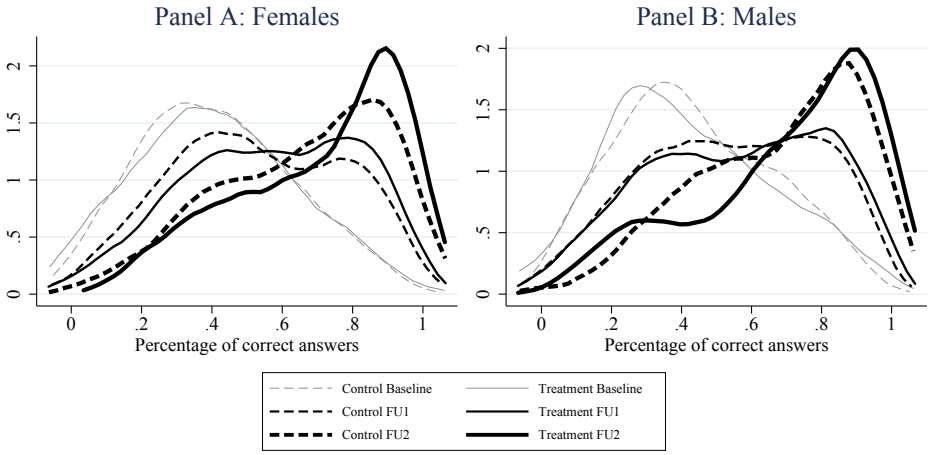


Figure A2: Distribution of Second/Third Grade Math Scores, by Gender

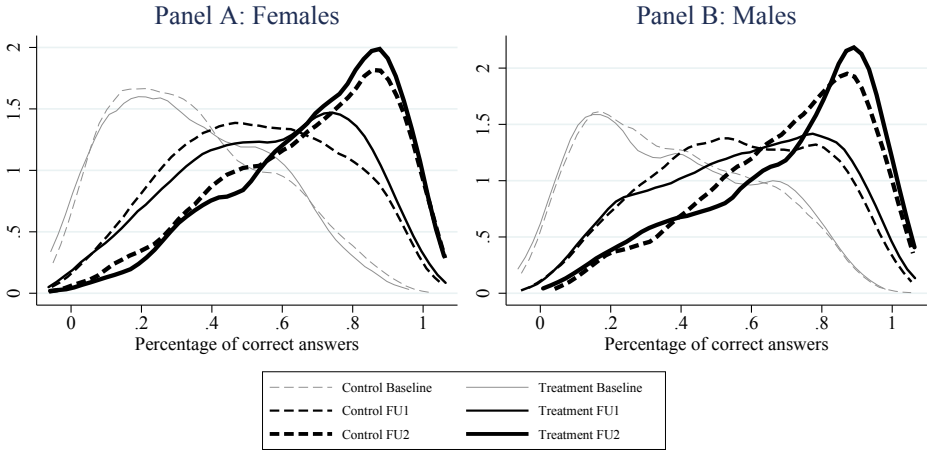
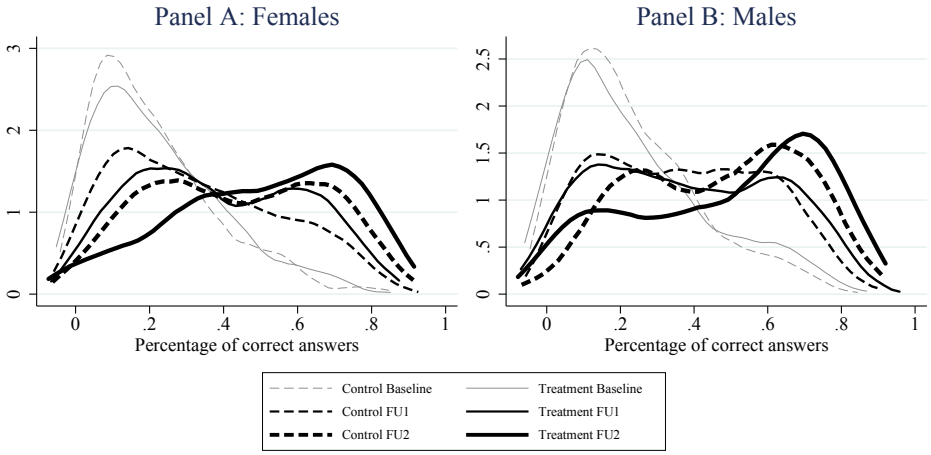


Figure A3: Distribution of Second/Third Grade Oral Scores, by Gender



Analagously, in Figures A4, A5, and A6, we display the densities of scores at first and second follow-up for boys and girls who were in fourth grade at baseline (and in fifth grade in the subsequent year). The differences between treatment and control schools are smaller than the ones we document for grade 3, especially when we consider girls. In the main body of the paper we show that there is no statistically significant impact of school grants on fifth grade test scores.

Figure A4: Distribution of Fourth/Fifth Grade French Scores, by Gender

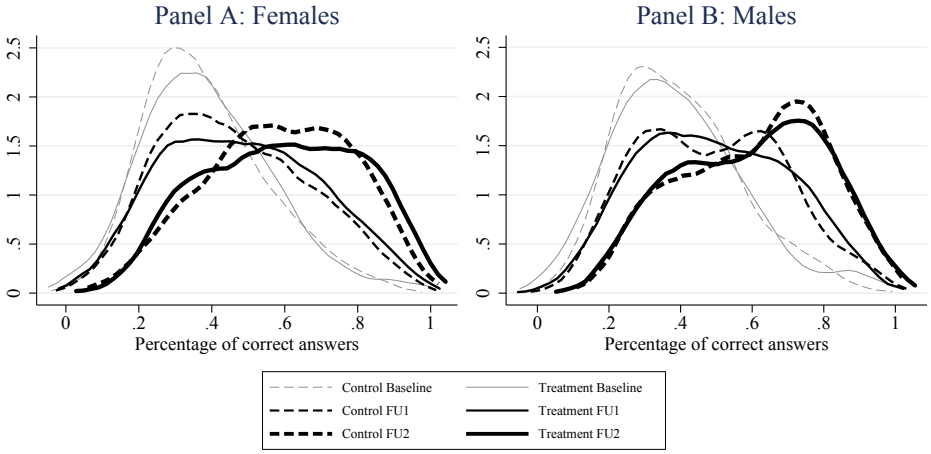


Figure A5: Distribution of Fourth/Fifth Grade Math Scores, by Gender

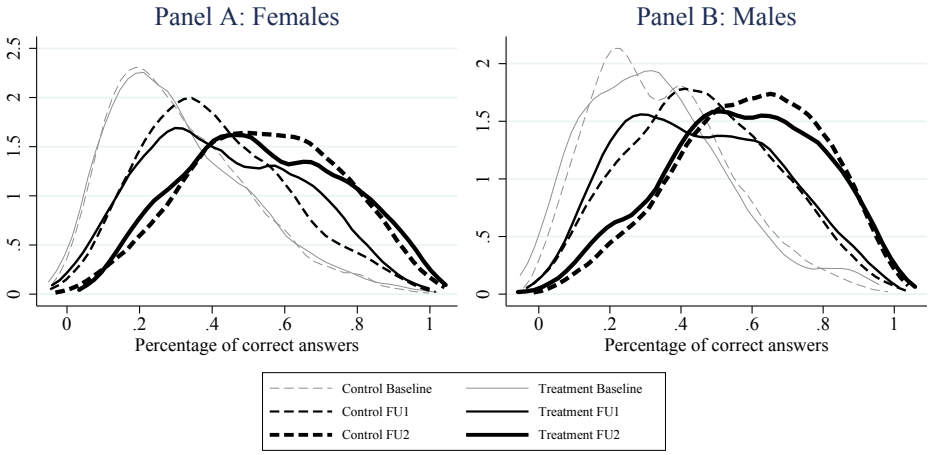
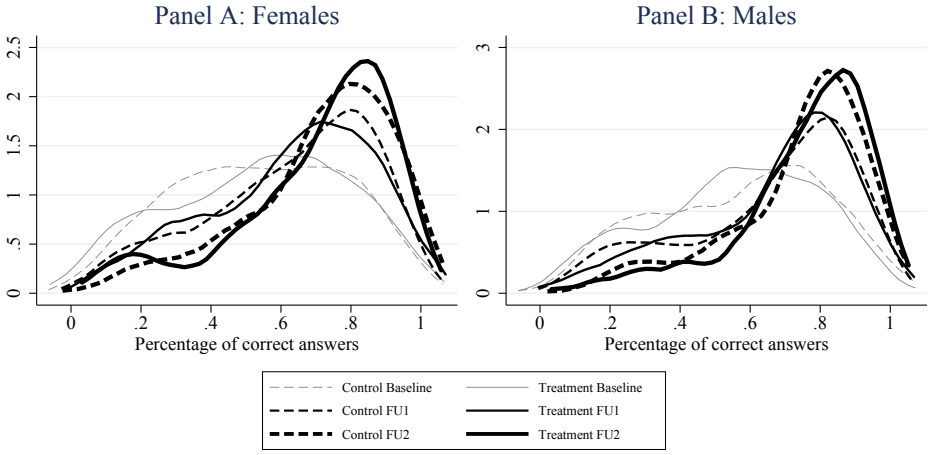
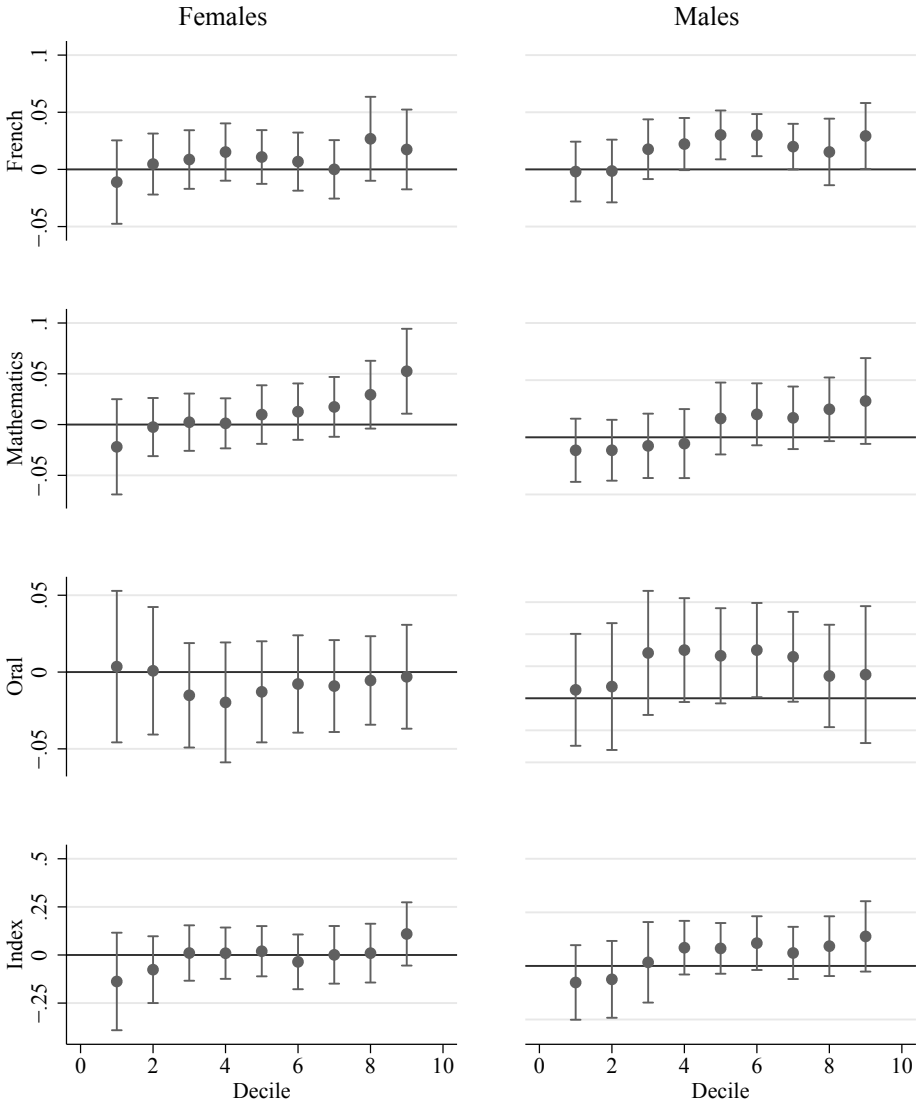


Figure A6: Distribution of Fourth/Fifth Grade Oral Scores, by Gender



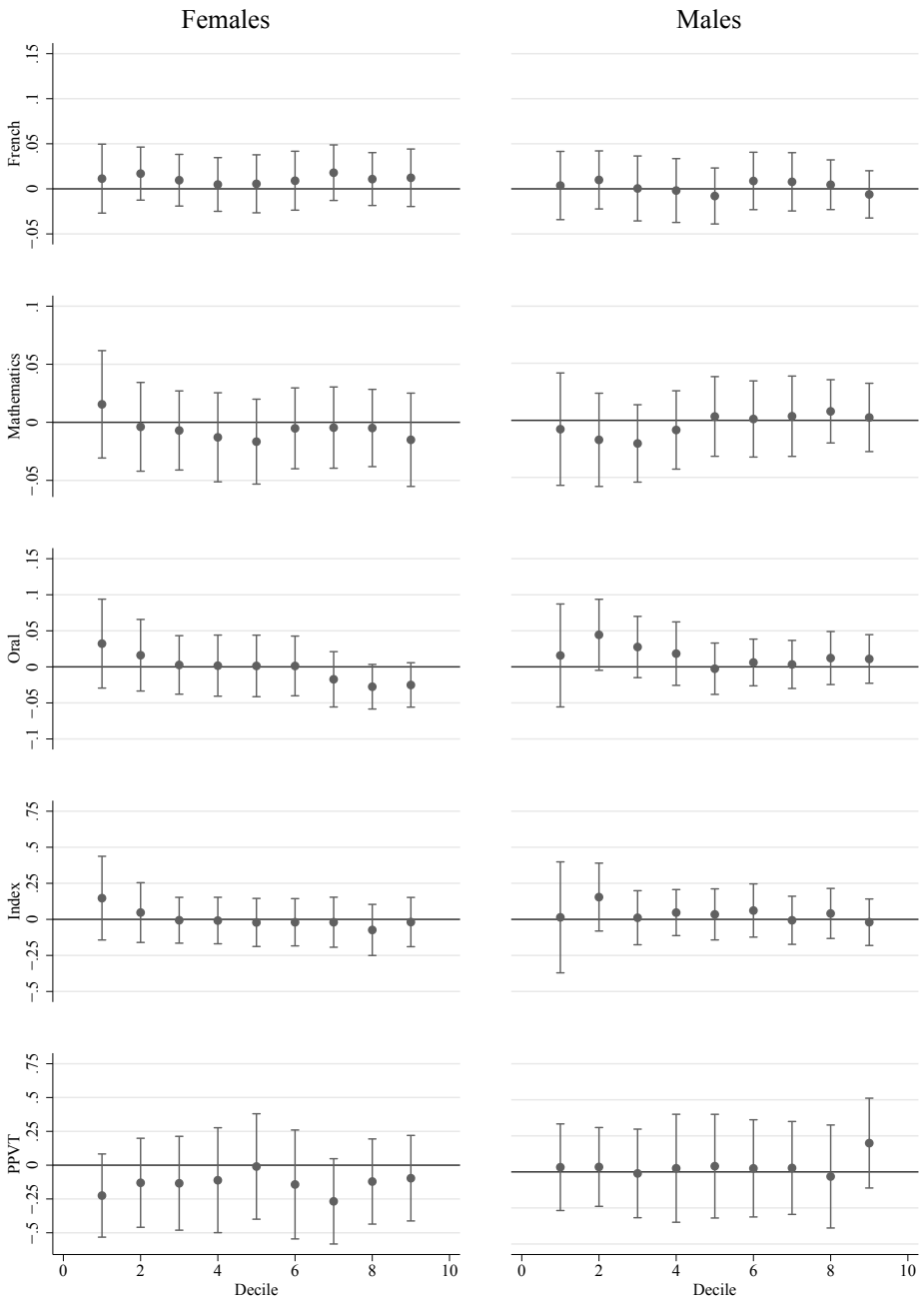
C. Other Appendix Tables and Figures

Figure A7: Distributional Impacts on Test Scores at the Beginning of Fifth Grade (First Follow-up)



Note: Point estimates from a quantile regression at each decile with 95% confidence intervals. Index coefficients are standardized.

Figure A8: Distributional Impacts on Test Scores at the End of Fifth Grade (Second Follow-up)



Note: Point estimates from a quantile regression at each decile with 95% confidence intervals. Index and PPVT plot coefficients are standardized.

Table A1: Baseline Descriptive Statistics and Balance (grade 4)

	Females		Males	
	Control	Treat-Control	Control	Treat-Control
Panel A: Test Scores				
Percent Correct: French	0.387	0.005	0.396	-0.004
	(0.168)	(0.015)	(0.171)	(0.016)
Percent Correct: Math	0.313	0.001	0.346	-0.013
	(0.185)	(0.017)	(0.187)	(0.017)
Percent Correct: Oral	0.537	-0.000	0.570	-0.016
	(0.241)	(0.024)	(0.244)	(0.022)
Index Score (standardized)	-0.061	-0.029	0.088	-0.063
	(0.970)	(0.103)	(0.989)	(0.096)
Panel B: Household Characteristics				
Days of school missed last week	0.155	-0.095*	0.153	-0.051
	(0.731)	(0.055)	(0.774)	(0.057)
Student works after school	0.018	-0.010	0.023	-0.017
	(0.132)	(0.011)	(0.149)	(0.010)
Household size	9.148	0.363	9.113	-0.107
	(3.951)	(0.466)	(4.172)	(0.402)
Number of children in household	5.155	0.522*	5.194	-0.081
	(2.671)	(0.300)	(2.883)	(0.284)
Head has any education	0.574	-0.127**	0.541	-0.008
	(0.495)	(0.054)	(0.499)	(0.050)
Percent of adult females with any education	0.313	-0.012	0.322	-0.013
	(0.396)	(0.041)	(0.392)	(0.039)
Distance to school (km)	0.646	0.052	0.628	0.132
	(0.745)	(0.098)	(0.768)	(0.082)
Parent involved in school	0.438	-0.092*	0.453	-0.051
	(0.497)	(0.052)	(0.499)	(0.049)
Expenditure on household food (1,000s CFA)	22.679	-0.190	21.689	1.270
	(16.038)	(1.686)	(15.390)	(1.459)
Expenditure on uniform (1,000s CFA)	2.360	0.178	2.259	-0.065
	(1.141)	(0.482)	(1.138)	(0.380)
Expenditure on tuition (1,000s CFA)	1.060	0.078	1.034	-0.008
	(1.036)	(0.124)	(1.006)	(0.090)
Expenditure on supplies (1,000s CFA)	4.328	-0.182	4.387	-0.573*
	(4.824)	(0.385)	(3.560)	(0.337)
Student has tutor	0.148	-0.066**	0.139	0.044
	(0.356)	(0.033)	(0.347)	(0.039)
Home has electricity	0.451	0.015	0.455	0.019
	(0.498)	(0.054)	(0.499)	(0.053)
Home has modern toilet	0.493	0.011	0.529	-0.002
	(0.501)	(0.056)	(0.500)	(0.051)
Land owned (hectares)	3.200	-0.585	2.605	-0.434
	(12.592)	(0.815)	(4.270)	(0.383)
Interview conducted in French	0.088	-0.013	0.129	-0.005
	(0.284)	(0.028)	(0.336)	(0.033)

Continued on next page

Table A1, continued: Baseline Descriptive Statistics and Balance (grade 4)

	All Students	
	Control	Treat-Control
Panel C: School and Teacher Characteristics		
Distance to nearest city (km)	18.033	0.213
	(24.555)	(2.197)
Locality population (100,000s)	1.410	0.031
	(4.428)	(0.453)
Locality has health center	0.714	0.030
	(0.452)	(0.043)
School located in South	0.187	-0.011
	(0.390)	(0.037)
School has Electricity	0.566	0.014
	(0.496)	(0.048)
Number of Teachers	9.741	0.575
	(4.927)	(0.517)
Number of Pupils	343.646	35.569
	(253.370)	(26.048)
School has library	0.210	0.082*
	(0.408)	(0.043)
Number of computers	1.303	0.014
	(4.390)	(0.404)
Number of manuals in classroom	66.431	5.679
	(51.957)	(5.405)
Percent teachers female	0.318	0.010
	(0.234)	(0.023)
Average teacher age	33.255	-0.097
	(4.233)	(0.390)
Percent of teachers with Baccalaureate	0.413	-0.024
	(0.223)	(0.022)
Average teacher experience	6.607	0.134
	(3.694)	(0.354)
Percent teachers with training in past 5 years	0.468	0.007
	(0.499)	(0.049)
Percent of principals with Baccalaureate	0.736	-0.058
	(0.441)	(0.045)

Columns 2 and 4 report the difference in means between treatment and control. Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: First Follow-up Descriptive Statistics and Balance (grade 3)

	Females		Males	
	Control	Treat-Control	Control	Treat-Control
Panel A: Household Characteristics				
Household size	10.026	0.127	9.891	0.135
	(4.586)	(0.466)	(4.264)	(0.440)
Number of children in household	5.481	0.354	5.459	0.112
	(2.852)	(0.299)	(2.773)	(0.273)
Head has any education	0.398	0.025	0.445	-0.010
	(0.490)	(0.048)	(0.498)	(0.050)
Percent adult females with any education	0.232	-0.020	0.216	0.037
	(0.347)	(0.034)	(0.329)	(0.037)
Distance to school (km)	0.640	-0.120**	0.593	-0.064
	(0.698)	(0.059)	(0.564)	(0.055)
Home has electricity	0.458	0.083	0.434	0.033
	(0.499)	(0.052)	(0.496)	(0.050)
Home has modern toilet	0.355	0.027	0.372	-0.060
	(0.479)	(0.050)	(0.484)	(0.047)
Land owned (hectares)	2.403	0.872	2.795	0.190
	(4.043)	(0.541)	(5.213)	(0.587)
Interview conducted in French	0.090	0.018	0.116	0.001
	(0.287)	(0.031)	(0.320)	(0.031)
	Everyone			
Panel B: School Characteristics				
Distance to nearest city (km)	18.347	0.097		
	(24.781)	(2.104)		
Population in locality (100,000s)	0.917	0.269		
	(2.744)	(0.313)		
Locality has health center	0.709	0.026		
	(0.454)	(0.043)		
School located in South	0.195	0.005		
	(0.396)	(0.037)		

Columns 2 and 4 report the difference in means between treatment and control. Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Summary Statistics of Nationally Representative Sample of Second and Fifth Grade Students

Panel A: School and Teacher Characteristics		
Locality has health center	0.809	(0.393)
School has electricity	0.359	(0.480)
Number of teachers	9.809	(5.007)
Number of students	500.683	(386.209)
School has library	0.217	(0.412)
Percent of teachers with Baccalaureate	0.474	(0.499)
Panel B: Household Characteristics		
Father literate	0.585	(0.493)
Mother literate	0.355	(0.478)
House has electricity	0.595	(0.491)
House has TV	0.598	(0.490)
House has modern toilet	0.367	(0.482)

Weighted means and standard deviations (in parentheses) shown. Source: PASEC 2006.

Table A4: Program Impacts on Grade 3 Test Scores, Standardized

Panel A: Beginning of Third Grade (First Follow-Up)

	French	Math	Oral
Female	0.149	0.128	0.210
	[0.022, 0.272]	[0.013, 0.247]	[0.061, 0.356]
Male	0.088	0.098	0.050
	[-0.053, 0.209]	[-0.029, 0.218]	[-0.102, 0.194]
Observations	2720	2718	1385

Panel B: End of Third Grade (Second Follow-Up)

	French	Math	Oral
Female	0.182	0.083	0.228
	[0.030, 0.344]	[-0.076, 0.236]	[0.041, 0.411]
Male	0.111	0.068	0.102
	[-0.032, 0.264]	[-0.086, 0.205]	[-0.081, 0.277]
Observations	1732	1721	853

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. Conditioning variables: Household size, number of children, education of head, distance to school, wealth index, interview language, baseline scores, missing dummies.

Table A5: Program Impacts on Grade 3 Test Scores, No Controls

Panel A: Beginning of Third Grade (First Follow-Up)

	French	Math	Oral	Index[a]
Female	0.041	0.031	0.055	0.240
	[0.004, 0.081]	[-0.009, 0.066]	[0.014, 0.097]	[0.054, 0.415]
Male	0.018	0.020	0.012	0.042
	[-0.019, 0.056]	[-0.017, 0.057]	[-0.029, 0.053]	[-0.137, 0.231]
Observations	2720	2718	1385	1350

Panel B: End of Third Grade (Second Follow-Up)

	French	Math	Oral	Index[a]	PPVT[a]
Female	0.046	0.021	0.065	0.290	0.134
	[-0.000, 0.091]	[-0.019, 0.062]	[0.013, 0.116]	[0.061, 0.529]	[-0.137, 0.386]
Male	0.020	0.008	0.025	0.068	0.198
	[-0.024, 0.061]	[-0.031, 0.047]	[-0.024, 0.076]	[-0.134, 0.280]	[-0.032, 0.436]
Observations	1732	1721	853	826	566

[a] The index and PPVT columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering.

Table A6: Program Impacts on Test Scores at Different Grades

	French		Math		Oral		Index[a]		PPVT[a]	
Grade	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Panel A: Beginning of Grade (First Follow-Up)										
3	0.041**	0.018	0.031*	0.020	0.055***	0.012	0.240**	0.042		
	(0.020)	(0.019)	(0.018)	(0.018)	(0.021)	(0.021)	(0.096)	(0.095)		
	[0.244]	[0.249]	[0.237]	[0.238]	[0.221]	[0.220]	[0.992]	[0.968]		
5	0.019	0.004	0.023	-0.003	-0.001	0.010	0.030	0.008		
	(0.017)	(0.016)	(0.017)	(0.017)	(0.022)	(0.022)	(0.097)	(0.097)		
	[0.201]	[0.203]	[0.198]	[0.207]	[0.235]	[0.240]	[0.956]	[1.018]		
6	0.016	0.016	0.024	0.014						
	(0.016)	(0.015)	(0.016)	(0.016)						
	[0.195]	[0.189]	[0.201]	[0.196]						
Panel B: End of Grade (Second Follow-Up)										
2	-0.010	0.013	-0.005	-0.000						
	(0.023)	(0.023)	(0.022)	(0.021)						
	[0.254]	[0.260]	[0.246]	[0.241]						
3	0.046**	0.020	0.021	0.008	0.065**	0.025	0.290	0.068	0.134	0.198*
	(0.022)	(0.021)	(0.020)	(0.021)	(0.027)	(0.026)	(0.118)	(0.109)	(0.131)	(0.117)
	[0.237]	[0.231]	[0.233]	[0.221]	[0.230]	[0.223]	[1.026]	[0.917]	[1.006]	[0.949]
4	-0.009	0.001	-0.021	0.003						
	(0.019)	(0.017)	(0.019)	(0.019)						
	[0.201]	[0.195]	[0.201]	[0.208]						
5	0.009	-0.005	-0.002	-0.020	-0.005	0.020	-0.046	0.051	-0.106	-0.046
	(0.020)	(0.018)	(0.020)	(0.019)	(0.024)	(0.021)	(0.115)	(0.109)	(0.128)	(0.125)
	[0.195]	[0.198]	[0.206]	[0.202]	[0.212]	[0.199]	[0.953]	[0.988]	[1.006]	[1.022]
6	0.004	0.007	0.000	-0.002						
	(0.017)	(0.016)	(0.018)	(0.016)						
	[0.183]	[0.174]	[0.193]	[0.169]						

[a] The index and PPVT columns are in standardized units.

Clustered standard errors in parentheses and standard deviation of control sample in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Program Impacts on Teacher and Household Outcomes, First Follow-Up

	Grade 3	Grade 5	Grade 3 - Grade 5
Panel A: Teacher Outcomes			
Teacher has Baccalaureate	-0.052	0.018	-0.070
	(0.047)	(0.046)	(0.064)
Teacher had training in past 5 years	0.083**	0.101**	-0.018
	(0.039)	(0.047)	(0.056)
Minutes spent preparing lesson	2.115	0.061	2.054
	(1.726)	(1.614)	(2.023)
Number of manuals	6.209	6.493	-0.284
	(4.231)	(4.964)	(5.296)
Number of measuring instruments	0.138	0.228	-0.090
	(0.179)	(0.193)	(0.210)
Number of chairs	0.018	0.071**	-0.053
	(0.038)	(0.035)	(0.041)
Teaches with books	0.004	0.012	-0.008
	(0.013)	(0.012)	(0.015)
Teaches with computers	-0.005	0.034	-0.038
	(0.019)	(0.024)	(0.029)
Times per day ask for silence	-1.638**	1.109	-2.747***
	(0.686)	(0.830)	(0.931)
Times per day punish a student	-0.126	0.741**	-0.868**
	(0.277)	(0.370)	(0.407)
Number of students who left in past year	0.303	1.063	-0.760
	(0.730)	(0.800)	(1.095)
Number of students who joined in past year	0.262	-0.030	0.292
	(0.212)	(0.185)	(0.194)
Panel B: Household Outcomes			
Days of school missed last week	0.106	0.010	0.096
	(0.075)	(0.055)	(0.086)
Student works after school	-0.008	-0.016	0.008
	(0.012)	(0.014)	(0.014)
Parent involved in school	0.037	0.026	0.011
	(0.038)	(0.038)	(0.039)
Expenditure on uniform (1,000s CFA)	-0.003	-0.070	0.067
	(0.070)	(0.056)	(0.055)
Expenditure on tuition (1,000s CFA)	0.303	0.035	0.269
	(0.303)	(0.164)	(0.313)
Expenditure on supplies (1,000s CFA)	-0.162	-0.111	-0.050
	(0.232)	(0.320)	(0.302)
Student has tutor	-0.027	-0.009	-0.017
	(0.026)	(0.029)	(0.031)
Expenditure on children (1,000s CFA)	0.303	-0.175	0.477
	(0.611)	(0.662)	(0.683)

Column 3 reports the difference in impacts between grade 3 and grade 5. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Characteristics of Schools by Baseline Missing Status, First Follow-Up

	Not Missing	Missing	Not Missing - Missing
Panel A: Treatment Status			
Treated	0.337	0.312	0.025
	(0.022)	(0.086)	(0.089)
Panel B: Control School Characteristics			
School located in South	0.186	0.439	-0.253**
	(0.022)	(0.114)	(0.116)
School located in Rural Area	0.736	1.000	-0.264***
	(0.025)	(0.000)	(0.025)
Locality Population (100,000s)	0.981	0.024	0.957***
	(0.173)	(0.006)	(0.173)
Number of Teachers	9.946	7.090	2.856***
	(0.278)	(0.951)	(0.991)
Number of Pupils	341.617	238.326	103.291***
	(14.283)	(35.026)	(37.826)
Percent Correct: French, Grade 3	0.540	0.418	0.123***
	(0.010)	(0.044)	(0.045)
Percent Correct: Math, Grade 3	0.550	0.408	0.142***
	(0.009)	(0.041)	(0.042)
Percent Correct: Oral, Grade 3	0.361	0.212	0.149***
	(0.010)	(0.036)	(0.037)
Percent Correct: Index, Grade 3	-0.012	-0.666	0.653***
	(0.047)	(0.197)	(0.203)
Percent Correct: French, Grade 5	0.483	0.379	0.104***
	(0.008)	(0.030)	(0.032)
Percent Correct: Math, Grade 5	0.441	0.342	0.099***
	(0.009)	(0.036)	(0.037)
Percent Correct: Oral, Grade 5	0.648	0.572	0.076
	(0.011)	(0.049)	(0.050)
Percent Correct: Index, Grade 5	0.026	-0.483	0.509***
	(0.045)	(0.173)	(0.179)

Column 3 reports the difference in means between not missing and missing schools. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Program Impacts on Grade 5 Test Scores by Baseline Ability

Panel A: Beginning of Grade 5 (First Follow-up)

	French		Math		Oral		Index[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
Low	0.001	0.017	-0.008	-0.010	-0.022	0.030	-0.047	0.045
	[-0.025, 0.031]	[-0.007, 0.046]	[-0.036, 0.022]	[-0.038, 0.017]	[-0.065, 0.018]	[-0.019, 0.081]	[-0.218, 0.110]	[-0.121, 0.221]
High	0.002	-0.001	0.020	0.009	0.003	0.029	-0.024	0.098
	[-0.039, 0.038]	[-0.039, 0.032]	[-0.020, 0.063]	[-0.024, 0.048]	[-0.029, 0.040]	[0.001, 0.058]	[-0.187, 0.128]	[-0.062, 0.256]
p-value (diff)	0.972	0.373	0.247	0.377	0.378	0.974	0.835	0.648

Panel B: End of Grade 5 (Second Follow-up)

	French		Math		Oral		Index[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
Low	-0.015	0.012	-0.025	-0.015	-0.062	0.010	-0.187	-0.017
	[-0.046, 0.015]	[-0.025, 0.051]	[-0.062, 0.014]	[-0.054, 0.027]	[-0.113, -0.006]	[-0.052, 0.074]	[-0.402, 0.025]	[-0.295, 0.246]
High	0.025	-0.010	0.014	-0.022	0.018	0.007	0.034	0.021
	[-0.008, 0.060]	[-0.043, 0.022]	[-0.027, 0.056]	[-0.063, 0.018]	[-0.019, 0.059]	[-0.031, 0.049]	[-0.152, 0.220]	[-0.165, 0.215]
p-value (diff)	0.082	0.363	0.150	0.791	0.020	0.930	0.103	0.815

[a] The index columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or Table 2 for control variables.The P-value row shows the p-value of the difference in the two treatment effects of the column.

Table A10: Program Impacts on Grade 5 Test Scores by Region

Panel A: Beginning of Grade 5 (First Follow-up)

	French		Math		Oral		Index[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
South	-0.032	-0.012	-0.005	-0.014	0.060	0.034	0.092	-0.057
	[-0.095, 0.029]	[-0.060, 0.037]	[-0.067, 0.055]	[-0.071, 0.055]	[-0.019, 0.142]	[-0.043, 0.116]	[-0.205, 0.385]	[-0.383, 0.263]
North	0.017	0.018	0.018	0.012	-0.021	0.018	-0.004	0.082
	[-0.007, 0.042]	[-0.006, 0.042]	[-0.010, 0.044]	[-0.015, 0.041]	[-0.053, 0.011]	[-0.013, 0.049]	[-0.130, 0.128]	[-0.044, 0.204]
p-value (diff)	0.162	0.303	0.511	0.454	0.073	0.698	0.554	0.443

Panel B: End of Grade 5 (Second Follow-up)

	French		Math		Oral		Index[a]		PPVT[a]	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
South	-0.019	0.014	-0.023	0.003	-0.001	0.053	0.010	0.163	-0.242	-0.188
	[-0.073, 0.047]	[-0.047, 0.067]	[-0.085, 0.035]	[-0.063, 0.057]	[-0.086, 0.088]	[-0.039, 0.139]	[-0.328, 0.318]	[-0.174, 0.544]	[-0.923, 0.540]	[-0.770, 0.416]
North	0.013	-0.006	0.003	-0.018	-0.006	0.014	-0.011	0.033	-0.097	0.061
	[-0.012, 0.041]	[-0.036, 0.027]	[-0.028, 0.030]	[-0.054, 0.013]	[-0.044, 0.031]	[-0.023, 0.054]	[-0.161, 0.134]	[-0.139, 0.206]	[-0.352, 0.205]	[-0.161, 0.311]
p-value (diff)	0.325	0.561	0.468	0.551	0.917	0.437	0.911	0.550	0.712	0.454

[a] The index and PPVT columns are in standardized units.

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or Table 2 for control variables.The P-value row shows the p-value of the difference in the two treatment effects of the column.

Table A11: Program Impacts on Grade 3 Test Scores by Region, Standardized

Panel A: Beginning of Grade 3 (First Follow-up)

	French		Math		Oral	
	Female	Male	Female	Male	Female	Male
South	0.505	0.314	0.288	0.375	0.456	0.215
	[0.231, 0.775]	[0.024, 0.580]	[0.025, 0.550]	[0.106, 0.641]	[0.199, 0.728]	[-0.152, 0.511]
North	0.060	0.038	0.089	0.032	0.156	0.013
	[-0.082, 0.203]	[-0.111, 0.185]	[-0.040, 0.222]	[-0.095, 0.164]	[-0.012, 0.338]	[-0.147, 0.168]

Panel B: End of Grade 3 (Second Follow-up)

	French		Math		Oral	
	Female	Male	Female	Male	Female	Male
South	0.369	0.282	0.287	0.417	0.606	0.227
	[-0.002, 0.695]	[-0.053, 0.672]	[-0.091, 0.615]	[0.057, 0.771]	[0.207, 0.992]	[-0.152, 0.584]
North	0.148	0.052	0.042	-0.034	0.130	0.061
	[-0.020, 0.334]	[-0.119, 0.222]	[-0.129, 0.202]	[-0.205, 0.135]	[-0.076, 0.350]	[-0.134, 0.260]

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or Table 2 for control variables.

Table A12: Baseline Descriptive Statistics and Balance (grade 2), South

	Females		Males	
	Control	Treat-Control	Control	Treat-Control
Panel A: Test Scores				
Percent Correct: French	0.431	-0.089*	0.489	-0.079
	(0.249)	(0.050)	(0.273)	(0.062)
Percent Correct: Math	0.304	-0.035	0.362	-0.012
	(0.225)	(0.046)	(0.230)	(0.055)
Percent Correct: Oral	0.114	0.056*	0.168	0.014
	(0.110)	(0.033)	(0.160)	(0.035)
Index Score (standardized)	-0.352	-0.073	-0.013	-0.291
	(0.895)	(0.256)	(0.999)	(0.250)
Panel B: Household Characteristics				
Days of school missed last week	0.158	0.101	0.041	-0.005
	(0.727)	(0.218)	(0.286)	(0.054)
Student works after school	0.000	0.074	0.000	0.000
	(0.000)	(0.049)	(0.000)	(0.000)
Household size	10.140	-0.103	9.918	-1.418*
	(4.116)	(1.112)	(4.242)	(0.768)
Number of children in household	5.754	-0.199	6.061	-0.990*
	(2.523)	(0.771)	(3.105)	(0.577)
Head has any education	0.648	-0.056	0.681	-0.125
	(0.482)	(0.112)	(0.471)	(0.121)
Percent of adult females with any education	0.324	-0.057	0.357	0.063
	(0.363)	(0.084)	(0.408)	(0.108)
Distance to school (km)	0.817	-0.274	0.976	-0.524***
	(1.045)	(0.176)	(1.100)	(0.179)
Parent involved in school	0.228	0.327***	0.265	0.020
	(0.423)	(0.117)	(0.446)	(0.106)
Expenditure on household food (1,000s CFA)	13.289	4.599	14.918	-1.613
	(8.821)	(2.909)	(9.403)	(2.092)
Expenditure on uniform (1,000s CFA)	4.500	-1.440*	3.000	0.000
	(0.866)	(0.694)	(0.500)	(0.776)
Expenditure on tuition (1,000s CFA)	1.368	-0.031	1.033	-0.072
	(1.493)	(0.352)	(0.862)	(0.194)
Expenditure on supplies (1,000s CFA)	2.618	-0.205	2.824	0.086
	(1.880)	(0.431)	(2.084)	(0.486)
Student has tutor	0.035	-0.035	0.041	0.031
	(0.186)	(0.025)	(0.200)	(0.057)
Home has electricity	0.281	-0.021	0.224	0.168
	(0.453)	(0.104)	(0.422)	(0.115)
Home has modern toilet	0.298	-0.187**	0.306	0.087
	(0.462)	(0.090)	(0.466)	(0.119)
Land owned (hectares)	1.965	2.628	2.551	-0.372
	(2.803)	(1.917)	(3.731)	(0.730)
Interview conducted in French	0.193	-0.045	0.265	-0.122
	(0.398)	(0.087)	(0.446)	(0.102)

Continued on next page

Table A12, continued: Baseline Descriptive Statistics and Balance (grade 2), South

	All Students	
	Control	Treat-Control
Panel C: School and Teacher Characteristics		
Distance to nearest city (km)	16.378	-0.873
	(16.536)	(3.097)
Locality population (100,000s)	0.139	-0.035
	(0.374)	(0.074)
Locality has health center	0.740	-0.010
	(0.439)	(0.100)
School has Electricity	0.497	-0.014
	(0.501)	(0.114)
Number of Teachers	10.348	-0.303
	(5.293)	(1.233)
Number of Pupils	366.652	-8.090
	(248.773)	(56.467)
School has library	0.069	0.099
	(0.254)	(0.077)
Number of computers	0.580	0.060
	(3.866)	(0.658)
Number of manuals in classroom	53.021	-5.602
	(42.262)	(9.259)
Percent teachers female	0.237	-0.013
	(0.221)	(0.045)
Average teacher age	32.932	-0.001
	(3.176)	(0.771)
Percent teachers with Baccalaureate	0.419	-0.049
	(0.214)	(0.051)
Average teacher experience	6.487	-0.345
	(2.570)	(0.572)
Percent teachers with training in past 5 years	0.370	0.089
	(0.483)	(0.115)
Percent of principals with Baccalaureate	0.730	-0.185*
	(0.444)	(0.110)

Columns 2 and 4 report the difference in means between treatment and control. Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Baseline Descriptive Statistics and Balance (grade 2), North

	Females		Males	
	Control	Treat-Control	Control	Treat-Control
Panel A: Test Scores				
Percent Correct: French	0.409	0.017	0.424	0.004
	(0.212)	(0.019)	(0.213)	(0.020)
Percent Correct: Math	0.360	-0.000	0.386	0.001
	(0.221)	(0.019)	(0.235)	(0.021)
Percent Correct: Oral	0.232	0.003	0.248	0.009
	(0.167)	(0.018)	(0.174)	(0.021)
Index Score (standardized)	-0.034	0.035	0.100	0.050
	(0.956)	(0.102)	(0.989)	(0.114)
Panel B: Household Characteristics				
Days of school missed last week	0.222	0.046	0.145	0.091
	(1.155)	(0.116)	(0.620)	(0.099)
Student works after school	0.012	0.003	0.012	-0.004
	(0.111)	(0.013)	(0.108)	(0.011)
Household size	9.182	0.310	9.183	0.003
	(4.019)	(0.456)	(4.028)	(0.501)
Number of children in household	5.083	0.277	5.214	-0.141
	(2.567)	(0.283)	(2.557)	(0.329)
Head has any education	0.595	-0.060	0.569	0.070
	(0.492)	(0.056)	(0.496)	(0.056)
Percent of adult females with any education	0.375	-0.036	0.364	-0.026
	(0.407)	(0.046)	(0.422)	(0.047)
Distance to school (km)	0.668	0.021	0.696	-0.020
	(0.757)	(0.085)	(0.959)	(0.097)
Parent involved in school	0.402	0.098*	0.427	0.003
	(0.491)	(0.056)	(0.496)	(0.057)
Expenditure on household food (1,000s CFA)	23.812	0.497	23.151	2.162
	(16.769)	(1.697)	(15.317)	(1.732)
Expenditure on uniform (1,000s CFA)	2.174	0.526	2.348	-0.410
	(1.221)	(0.568)	(0.994)	(0.389)
Expenditure on tuition (1,000s CFA)	1.054	-0.008	1.121	0.023
	(0.966)	(0.113)	(1.345)	(0.136)
Expenditure on supplies (1,000s CFA)	3.659	0.204	4.559	-1.078*
	(2.558)	(0.304)	(8.232)	(0.585)
Student has tutor	0.174	0.013	0.180	-0.035
	(0.380)	(0.045)	(0.385)	(0.041)
Home has electricity	0.525	0.046	0.518	-0.009
	(0.500)	(0.058)	(0.501)	(0.057)
Home has modern toilet	0.583	0.019	0.623	-0.042
	(0.494)	(0.057)	(0.486)	(0.055)
Land owned (hectares)	2.298	-0.244	2.499	0.840
	(3.173)	(0.357)	(3.808)	(0.982)
Interview conducted in French	0.095	-0.025	0.093	-0.037
	(0.294)	(0.029)	(0.292)	(0.031)

Continued on next page

Table A13, continued: Baseline Descriptive Statistics and Balance (grade 2), North

	All Students	
	Control	Treat-Control
Panel C: School and Teacher Characteristics		
Distance to nearest city (km)	18.830	0.084
	(26.534)	(2.562)
Locality population (100,000s)	1.666	0.039
	(4.828)	(0.556)
Locality has health center	0.703	0.034
	(0.457)	(0.048)
School has Electricity	0.582	0.018
	(0.493)	(0.052)
Number of Teachers	9.527	0.606
	(4.878)	(0.562)
Number of Pupils	335.334	36.625
	(252.917)	(28.665)
School has library	0.237	0.073
	(0.426)	(0.048)
Number of computers	1.440	-0.034
	(4.481)	(0.461)
Number of manuals in classroom	61.414	5.074
	(45.676)	(5.153)
Percent teachers female	0.334	0.014
	(0.234)	(0.025)
Average teacher age	33.160	-0.162
	(4.446)	(0.444)
Percent teachers with Baccalaureate	0.411	-0.013
	(0.229)	(0.025)
Average teacher experience	6.574	0.171
	(3.901)	(0.406)
Percent teachers with training in past 5 years	0.496	0.097*
	(0.500)	(0.054)
Percent of principals with Baccalaureate	0.736	-0.021
	(0.441)	(0.048)

Columns 2 and 4 report the difference in means between treatment and control. Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A14: Program Impacts on Grade 3 Household Characteristics by Region, First Follow-Up

	Female			Male		
	South	North	South-North	South	North	South-North
Student works after school	-0.016	-0.033*	0.017	0.047	0.006	0.041
	(0.016)	(0.019)	(0.025)	(0.050)	(0.019)	(0.053)
Days of school missed last week	0.498*	0.107	0.391	0.195	0.009	0.185
	(0.277)	(0.087)	(0.291)	(0.167)	(0.133)	(0.214)
Parent involved in school	0.026	0.043	-0.017	-0.094	0.061	-0.155
	(0.102)	(0.058)	(0.117)	(0.101)	(0.055)	(0.115)
Expenditure on uniform (1,000s CFA)	0.011	0.026	-0.015	-0.034	-0.039	0.005
	(0.240)	(0.083)	(0.254)	(0.335)	(0.054)	(0.339)
Expenditure on tuition (1,000s CFA)	1.294	0.064	1.230	1.404	0.059	1.345
	(1.418)	(0.158)	(1.427)	(1.509)	(0.132)	(1.515)
Expenditure on supplies (1,000s CFA)	-0.042	-0.026	-0.015	-0.040	-0.221	0.181
	(0.532)	(0.356)	(0.640)	(0.561)	(0.342)	(0.658)
Student has tutor	-0.032	-0.046	0.014	-0.044	-0.013	-0.031
	(0.022)	(0.042)	(0.048)	(0.049)	(0.042)	(0.064)

Columns 3 and 6 report the difference in program impacts between South and North. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A15: Student Test Score Sample Sizes and Attrition

	Grade 3			Grade 5		
	French	Math	Oral	French	Math	Oral
Baseline Sample Size	2722	2752	1388	2724	2726	1362
First Followup Sample Size	2720	2718	1385	2648	2643	1347
New Observations	322	299	177	262	261	155
Total Attrition	324	333	180	338	344	170
% Attrition	0.119	0.121	0.130	0.124	0.126	0.125
% Attrition, Treated	0.118	0.117	0.113	0.116	0.116	0.114
% Attrition, Control	0.119	0.123	0.138	0.128	0.131	0.130
Second Followup Sample Size	1732	1721	853	1606	1606	833
Total Attrition*	290	301	208	355	357	186
% Attrition*	0.157	0.162	0.222	0.197	0.197	0.206
% Attrition*, Treated	0.160	0.165	0.230	0.204	0.206	0.224
% Attrition*, Control	0.155	0.159	0.215	0.189	0.189	0.188
Observed in All Waves	1464	1461	709	1396	1392	696

*Attrition in the second follow-up is based on cohorts 1 and 3, since cohort 2 schools were dropped in the second follow-up. A student in the second follow-up has attrited if they have a baseline test score but not a second follow-up test score (regardless of their status in the first follow-up).

Table A16: Difference (Treat-Control) in Baseline Characteristics Among First Follow-up Non-Attriters

	Grade 2		Grade 4	
	Females	Males	Females	Males
Panel A: Test Scores				
Percent Correct: French	0.003	-0.013	0.010	-0.009
	(0.019)	(0.021)	(0.016)	(0.015)
Percent Correct: Math	-0.001	-0.009	0.012	-0.016
	(0.019)	(0.020)	(0.018)	(0.017)
Percent Correct: Oral	0.019	0.007	0.001	-0.012
	(0.017)	(0.019)	(0.025)	(0.022)
Index Score (standardized)	0.028	-0.056	-0.014	-0.068
	(0.105)	(0.113)	(0.100)	(0.094)
Panel B: Household Characteristics				
Days of school missed last week	0.074	0.078	-0.100*	-0.053
	(0.114)	(0.092)	(0.058)	(0.063)
Student works after school	0.017	-0.000	-0.012	-0.015
	(0.016)	(0.009)	(0.012)	(0.011)
Household size	0.069	-0.358	0.439	-0.045
	(0.432)	(0.474)	(0.512)	(0.419)
Number of children in household	0.045	-0.276	0.579*	-0.074
	(0.281)	(0.315)	(0.329)	(0.304)
Head has any education	-0.095*	0.046	-0.132**	-0.035
	(0.053)	(0.054)	(0.058)	(0.053)
Percent of adult females with any education	-0.020	-0.004	0.006	-0.026
	(0.044)	(0.046)	(0.043)	(0.042)
Distance to school (km)	-0.051	-0.131	0.062	0.165*
	(0.084)	(0.094)	(0.108)	(0.092)
Parent involved in school	0.114**	-0.009	-0.070	-0.070
	(0.053)	(0.053)	(0.055)	(0.052)
Expenditure on household food (1,000s CFA)	2.122	-0.121	0.324	2.192
	(1.670)	(1.572)	(1.850)	(1.568)
Expenditure on uniform (1,000s CFA)	0.543	-0.271	0.064	-0.313
	(0.528)	(0.537)	(0.507)	(0.368)
Expenditure on tuition (1,000s CFA)	0.066	-0.011	0.016	0.001
	(0.111)	(0.128)	(0.123)	(0.097)
Expenditure on supplies (1,000s CFA)	-0.047	-0.986*	-0.011	-0.693*
	(0.254)	(0.565)	(0.416)	(0.371)
Student has tutor	0.008	-0.049	-0.055	0.060
	(0.038)	(0.037)	(0.035)	(0.042)
Home has electricity	0.067	0.016	0.081	0.024
	(0.055)	(0.055)	(0.057)	(0.055)
Home has modern toilet	0.020	-0.047	0.045	-0.017
	(0.055)	(0.054)	(0.059)	(0.054)
Land owned (hectares)	-0.045	0.661	-0.120	-0.310
	(0.469)	(0.915)	(0.466)	(0.421)
Interview conducted in French	-0.026	-0.048	-0.002	-0.025
	(0.030)	(0.033)	(0.031)	(0.034)

Continued on next page

Table A16, continued: Difference (Treat-Control) in Baseline Characteristics Among First Follow-up Non-Attriters

	Grade 2 (All students)	Grade 4 (All Students)
Panel C: School and Teacher Characteristics		
Distance to nearest city (km)	0.347	0.035
	(2.203)	(2.341)
Locality population (100,000s)	-0.063	0.142
	(0.407)	(0.473)
Locality has health center	0.032	0.032
	(0.044)	(0.043)
School located in South	-0.016	-0.021
	(0.038)	(0.037)
School has Electricity	0.021	0.037
	(0.049)	(0.049)
Number of Teachers	0.464	0.534
	(0.515)	(0.515)
Number of Pupils	33.657	31.181
	(25.444)	(26.044)
School has library	0.085**	0.094**
	(0.043)	(0.044)
Number of computers	-0.000	0.052
	(0.397)	(0.409)
Number of manuals in classroom	3.810	6.530
	(4.564)	(5.498)
Percent teachers female	0.007	0.012
	(0.023)	(0.023)
Average teacher age	-0.126	0.051
	(0.394)	(0.392)
Percent teachers with Baccalaureate	-0.012	-0.017
	(0.023)	(0.023)
Average teacher experience	0.085	0.220
	(0.345)	(0.355)
Percent teachers with training in past 5 years	0.092*	0.012
	(0.050)	(0.050)
Percent principals with Baccalaureate	-0.034	-0.037
	(0.045)	(0.045)

Point estimates of the difference in characteristic between treated students and non-treated students, among students who did not attrit between baseline and first follow-up. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A17: Difference (Treat-Control) in Baseline Characteristics Among Second Follow-up Non-Attriters

	Grade 2		Grade 4	
	Females	Males	Females	Males
Panel A: Test Scores				
Percent Correct: French	0.021	-0.009	0.004	-0.001
	(0.023)	(0.025)	(0.018)	(0.018)
Percent Correct: Math	0.011	-0.011	0.004	-0.011
	(0.022)	(0.024)	(0.020)	(0.020)
Percent Correct: Oral	0.020	0.001	-0.006	-0.004
	(0.019)	(0.023)	(0.029)	(0.027)
Index Score (standardized)	0.104	0.008	-0.049	-0.005
	(0.128)	(0.133)	(0.116)	(0.110)
Panel B: Household Characteristics				
Days of school missed last week	0.043	0.064	-0.126*	-0.152
	(0.169)	(0.111)	(0.076)	(0.097)
Student works after school	0.026	-0.007	-0.015	-0.031
	(0.018)	(0.013)	(0.017)	(0.019)
Household size	0.551	0.339	0.426	-0.399
	(0.513)	(0.538)	(0.596)	(0.529)
Number of children in household	0.207	-0.027	0.466	-0.249
	(0.341)	(0.366)	(0.404)	(0.375)
Head has any education	-0.074	0.041	-0.105	-0.036
	(0.064)	(0.063)	(0.069)	(0.064)
Percent of adult females with any education	-0.025	0.017	0.038	-0.035
	(0.055)	(0.055)	(0.053)	(0.049)
Distance to school (km)	-0.097	-0.163	0.045	0.206**
	(0.116)	(0.103)	(0.117)	(0.102)
Parent involved in school	0.097	0.021	-0.041	-0.087
	(0.065)	(0.063)	(0.066)	(0.063)
Expenditure on household food (1,000s CFA)	3.075	1.924	-0.047	0.751
	(2.020)	(1.853)	(2.281)	(2.110)
Expenditure on uniform (1,000s CFA)	0.297	-0.438	-0.122	-0.500
	(0.736)	(0.644)	(0.693)	(0.442)
Expenditure on tuition (1,000s CFA)	0.053	-0.129	-0.092	-0.126
	(0.135)	(0.177)	(0.170)	(0.146)
Expenditure on supplies (1,000s CFA)	0.129	-1.238	-0.650	-0.425
	(0.355)	(0.982)	(0.670)	(0.445)
Student has tutor	0.020	-0.028	-0.064	0.049
	(0.041)	(0.043)	(0.043)	(0.046)
Home has electricity	0.070	-0.004	0.067	0.019
	(0.067)	(0.066)	(0.070)	(0.068)
Home has modern toilet	-0.052	-0.006	0.054	-0.065
	(0.066)	(0.064)	(0.069)	(0.066)
Land owned (hectares)	0.257	1.246	0.206	-0.481
	(0.550)	(1.009)	(0.462)	(0.575)
Interview conducted in French	-0.020	-0.038	-0.006	0.010
	(0.039)	(0.041)	(0.036)	(0.041)

Continued on next page

Table A17, continued: Difference (Treat-Control) in Baseline Characteristics Among Second Follow-up Non-Attriters

	Grade 2 (All students)	Grade 4 (All Students)
Panel C: School and Teacher Characteristics		
Distance to nearest city (km)	-0.513	-0.985
	(2.436)	(2.579)
Locality population (100,000s)	0.160	0.295
	(0.442)	(0.500)
Locality has health center	0.076	0.069
	(0.052)	(0.052)
School located in South	-0.024	-0.040
	(0.046)	(0.044)
School has Electricity	-0.019	0.003
	(0.056)	(0.057)
Number of Teachers	0.636	0.708
	(0.587)	(0.609)
Number of Pupils	43.355	47.491
	(28.603)	(30.330)
School has library	0.087*	0.079
	(0.047)	(0.049)
Number of computers	-0.144	-0.170
	(0.485)	(0.515)
Number of manuals in classroom	3.572	10.830*
	(5.758)	(6.374)
Percent teachers female	0.028	0.032
	(0.026)	(0.026)
Average teacher age	-0.056	0.020
	(0.459)	(0.454)
Percent teachers with Baccalaureate	-0.016	-0.017
	(0.027)	(0.027)
Average teacher experience	0.138	0.168
	(0.394)	(0.404)
Percent teachers with training in past 5 years	0.071	0.024
	(0.058)	(0.058)
Percent principals with Baccalaureate	-0.022	-0.026
	(0.052)	(0.053)

Point estimates of the difference in characteristic between treated students and non-treated students, among students who did not attrit between baseline and first follow-up. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

ANSD (2007). La pauvrete au senegal: de la devaluation de 1994 a 2001-2002. Technical report, Agence Nationale de la Statistique et de la Demographie.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Blimpo, M., Evans, D., and Lahire, N. (2014). School-based management and educational outcomes, lessons from a randomized field experiment.

Bruns, B., Filmer, D., and Patrinos, H. A. (2011). *Making schools work: New evidence on accountability reforms*. World Bank Publications.

Galiani, S., Gertler, P., and Schargrodsky, E. (2008). School decentralization: Helping the good get better, but leaving the poor behind. *Journal of Public Economics*, 92(10):2106–2120.

Galiani, S. and Perez-Truglia, R. (2013). School management in developing countries. *Education Policy in Developing Countries*, 4:193.

Glewwe, P., Hanushek, E. A., Humpage, S., and Ravina, R. (2013). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. *Education Policy in Developing Countries*, 4:13.

Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2:945–1017.

Glewwe, P., Kremer, M., and Moulin, S. (2009). Many children left behind? textbooks and test scores in kenya. *American Economic Journal: Applied Economics*, 1(1):112–135.

Glewwe, P., Kremer, M., Moulin, S., and Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in kenya. *Journal of Development Economics*, 74(1):251–268.

Hanushek, E. A. (2006). School resources. *Handbook of the Economics of Education*, 2:865–908.

Hanushek, E. A., Link, S., and Woessmann, L. (2013). Does school autonomy make sense everywhere? panel estimates from pisa. *Journal of Development Economics*, 104:212–232.

Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, 2:1051–1078.

Hanushek, E. A. and Woessmann, L. (2010). The economics of international differences in educational achievement. Technical report, National Bureau of Economic Research.

Murnane, R. J. and Ganimian, A. J. (2014). Improving educational outcomes in developing countries: Lessons from rigorous evaluations. Technical report, National Bureau of Economic Research.

PASEC (2007). Evaluation pasec senegal. Technical report, Program on the Analysis of Education Systems.

Reinikka, R. and Svensson, J. (2004). Local capture: evidence from a central government transfer program in uganda. *The Quarterly Journal of Economics*, 119(2):679–705.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

World Bank (2013). Projet d'amelioration de la qualite et de l'equite de l'education de base. Technical report.