

# School Grants and Education Quality

## Experimental Evidence from Senegal

*Pedro Carneiro*  
*Oswald Koussihouède*  
*Nathalie Lahire*  
*Costas Meghir*  
*Corina Mommaerts*



**WORLD BANK GROUP**

Education Global Practice Group

April 2016

## Abstract

The effect of increasing school resources on educational outcomes is a central issue in the debate on improving school quality. This paper uses a randomized experiment to analyze the impact of a school grants program in Senegal, which allowed schools to apply for funding for improvements of their own choice. The analysis finds positive

effects on test scores at lower grades that persist at least two years. These effects are concentrated among schools that focused funds on human resource improvements rather than school materials, suggesting that teachers and principals may be a central determinant of school quality.

---

This paper is a product of the Education Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [nlahire@worldbank.org](mailto:nlahire@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# School Grants and Education Quality: Experimental Evidence from Senegal

Pedro Carneiro      Oswald Koussihouèdé      Nathalie Lahire      Costas Meghir  
Corina Mommaerts\*

*JEL classification:* H52; I22; I25; O15

*Keywords:* Quality of education; Decentralization; School resources; Child Development; Clustered Randomized Control Trials

---

\*This is a revised version of “Decentralizing Education Resources: School Grants in Senegal”. Pedro Carneiro is at University College London, IFS and CEMMAP. Oswald Koussihouèdé is at the “Programme for the Analysis of Education Systems of the “Conférence des Ministres de l’Education des Etats et gouvernements de la Francophonie” (PASEC). Nathalie Lahire is at the World Bank. Corina Mommaerts and Costas Meghir are at Yale University. We thank David Evans, Deon Filmer and Waly Wane for helpful comments. Financial support was provided by the Education Program Development Fund (EPDF) of the World Bank. We thank Ibrahima Mbengue and officials from the ministry of education as well as the research team at the “Université Gaston Berger - Saint-Louis du Sénégal” for invaluable help during the field work. 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.

# 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).

The approach used in Senegal is one of decentralization of school resources, in the sense that it is the schools themselves that define their needs. 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)). We cannot conclude whether the grants approach is superior to an alternative where resources are directed centrally, because no other approach was tried. However, 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.<sup>1</sup> Almost all classroom instruction is conducted

---

<sup>1</sup>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)).

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.

## **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'Éducation 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.

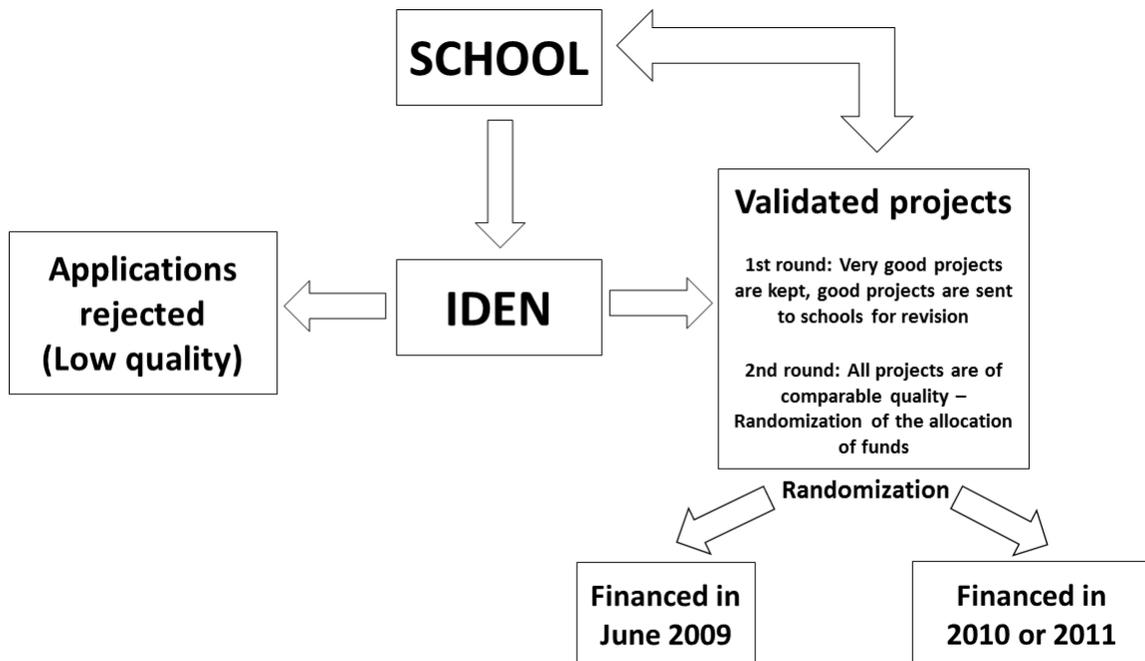
Generally, the program worked as follows. The Ministry of Education issued a call for proposals, based on the available grant funding, priority areas, and eligible activities (and sometimes eligible regions). Schools that decided to apply for funding completed a grant application for a school project (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).<sup>2</sup> We next 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.

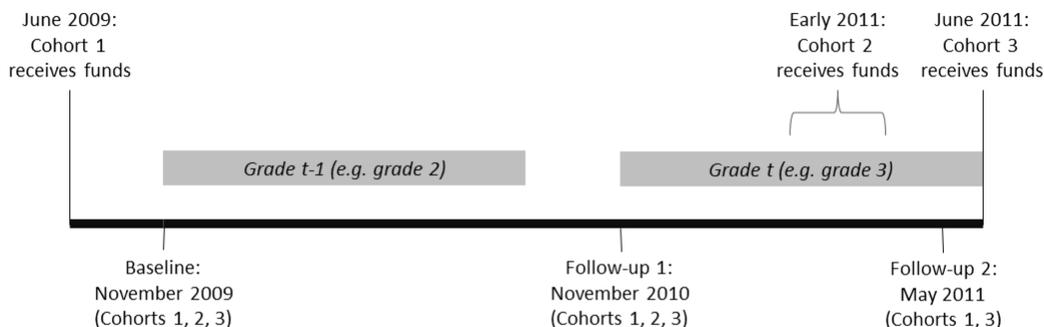
Figure 1: Evaluation Design



To implement this process, the procedures manual for the projets d’école was amended (relative to versions used for earlier cohorts of school grants) to include the revision of strong proposals needing adjustments. An additional official document issued by the Ministry of Education was circulated throughout the IDENs in the country, establishing the procedure described above as the norm for the allocation of funds for the next cohort of school projects.

<sup>2</sup>These numbers are based on collected self-reports from principals and teachers in our sample.

Figure 2: Evaluation Timeline



This process resulted in the selection of 633 projects to fund, whose locations are shown in Figure 4 in Appendix B.<sup>3</sup> 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 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 2).

<sup>3</sup>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.

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. 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.<sup>4</sup>

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<sup>5</sup> (among which 517 were successfully surveyed), and 340 at second follow-up (among which 325 were successfully surveyed and tested).<sup>6</sup> 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 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.

Table 1 shows descriptive statistics and balance between treatment and control schools for grades 2 and 4. Columns 2 and 4 show means and standard deviations of baseline characteristics in control schools, and columns 3 and 5 report the differences in characteristics between treatment and control at baseline and their standard errors. Panel A reports test scores.<sup>7</sup> The resulting mean scores for the French, mathematics, and oral tests (calculated as the proportion of correct responses on the exam) were around 20-40 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

---

<sup>4</sup>During the 2nd follow up we also tested a random sample of 2nd, 4th, and 6th grade students in French and mathematics, but did not collect any household information on them. However, in this study we concentrate on the panel of children we originally selected at baseline as planned in the randomization protocol. This ensures our results are not in any way affected by composition effects due to mobility of children that could have been induced as a result of the program.

<sup>5</sup>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.

<sup>6</sup>See Appendix Table 14 for the corresponding number of student-level observations and attrition. In Appendix Table 15 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).

<sup>7</sup>The full distribution of test scores is in Appendix A.

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.

Treatment and control schools are very well balanced. All but two differences (parental involvement in school and the percent of teachers who report receiving training in the past five years, both for second grade) are insignificant at the 5% level. 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.<sup>8</sup>

As 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 6 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 time-invariant 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 7, we show characteristics of a nationally representative sample of Senegalese households using data collected

---

<sup>8</sup>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 in Appendix B that the distribution of scores is highly non-normal, so a one standard deviation in test scores does not have the usual meaning. Nevertheless, for our main results we report standard deviations of control schools to convert results to standard deviations.

Table 1: Baseline Descriptive Statistics and Balance, by Grade

	Grade 2				Grade 4			
	Control		Treat-Control		Control		Treat-Control	
<b>Panel A: Test Scores</b>								
Percent Correct: French	0.42	(0.22)	-0.01	(0.02)	0.39	(0.17)	0.00	(0.01)
Percent Correct: Math	0.37	(0.23)	-0.00	(0.02)	0.33	(0.19)	-0.01	(0.02)
Percent Correct: Oral	0.22	(0.17)	0.01	(0.02)	0.55	(0.24)	-0.01	(0.02)
Index Score (standardized)	0.00	(0.98)	-0.00	(0.09)	0.02	(0.98)	-0.05	(0.09)
<b>Panel B: Household Characteristics</b>								
Days of school missed last week	0.17	(0.86)	0.07	(0.07)	0.16	(0.75)	-0.07*	(0.04)
Student works after school	0.01	(0.10)	0.01	(0.01)	0.02	(0.14)	-0.01*	(0.01)
Household size	9.26	(4.06)	0.00	(0.32)	9.07	(4.05)	0.14	(0.33)
Number of children in household	5.25	(2.61)	-0.03	(0.20)	5.14	(2.77)	0.21	(0.22)
Head has any education	0.60	(0.49)	-0.02	(0.04)	0.56	(0.50)	-0.06	(0.04)
Percent of adult females with any education	0.37	(0.41)	-0.03	(0.03)	0.31	(0.39)	-0.01	(0.03)
Distance to school (km)	0.71	(0.91)	-0.07	(0.06)	0.76	(2.56)	-0.03	(0.12)
Parent involved in school	0.38	(0.49)	0.09**	(0.04)	0.45	(0.50)	-0.07*	(0.04)
Expenditure on household food (1,000s CFA)	21.83	(15.45)	1.26	(1.17)	22.11	(15.50)	0.51	(1.20)
Expenditure on uniform (1,000s CFA)	2.43	(1.19)	0.15	(0.36)	2.28	(1.13)	0.06	(0.35)
Expenditure on tuition (1,000s CFA)	1.10	(1.18)	-0.01	(0.09)	1.03	(1.01)	0.04	(0.09)
Expenditure on supplies (1,000s CFA)	3.85	(5.64)	-0.38	(0.29)	4.34	(4.16)	-0.38	(0.27)
Student has tutor	0.15	(0.36)	-0.01	(0.03)	0.14	(0.35)	-0.01	(0.03)
Home has electricity	0.47	(0.50)	0.03	(0.04)	0.45	(0.50)	0.02	(0.04)
Home has modern toilet	0.54	(0.50)	-0.01	(0.04)	0.50	(0.50)	0.01	(0.04)
Land owned (hectares)	2.37	(3.47)	0.43	(0.46)	2.89	(9.11)	-0.50	(0.43)
Interview conducted in French	0.12	(0.32)	-0.04*	(0.02)	0.11	(0.31)	-0.01	(0.02)
<b>Panel C: School and Teacher Characteristics</b>								
Distance to nearest city (km)	18.38	(25.01)	-0.07	(2.18)	18.03	(24.56)	0.21	(2.20)
Locality population (100,000s)	1.38	(4.40)	0.04	(0.46)	1.41	(4.43)	0.03	(0.45)
Locality has health center	0.71	(0.45)	0.03	(0.04)	0.71	(0.45)	0.03	(0.04)
School located in South	0.18	(0.39)	-0.01	(0.04)	0.19	(0.39)	-0.01	(0.04)
School has Electricity	0.57	(0.50)	0.01	(0.05)	0.57	(0.50)	0.01	(0.05)
Number of Teachers	9.68	(4.97)	0.44	(0.51)	9.74	(4.93)	0.57	(0.52)
Number of Pupils	341.11	(252.39)	28.47	(25.60)	343.65	(253.37)	35.57	(26.05)
School has library	0.21	(0.40)	0.08*	(0.04)	0.21	(0.41)	0.08*	(0.04)
Number of computers	1.28	(4.39)	-0.01	(0.40)	1.30	(4.39)	0.01	(0.40)
Number of manuals in classroom	59.90	(45.18)	3.17	(4.58)	66.43	(51.96)	5.68	(5.40)
Percent teachers female	0.32	(0.24)	0.01	(0.02)	0.32	(0.23)	0.01	(0.02)
Average teacher age	33.12	(4.24)	-0.13	(0.39)	33.26	(4.23)	-0.10	(0.39)
Percent of teachers with Baccalaureate	0.41	(0.23)	-0.02	(0.02)	0.41	(0.22)	-0.02	(0.02)
Average teacher experience	6.56	(3.69)	0.08	(0.35)	6.61	(3.69)	0.13	(0.35)
Percent teachers with training in past 5 years	0.47	(0.50)	0.10**	(0.05)	0.47	(0.50)	0.01	(0.05)
Percent of principals with Baccalaureate	0.74	(0.44)	-0.05	(0.04)	0.74	(0.44)	-0.06	(0.04)

*Notes:* Grouped columns 2 and 4 report means and standard deviations of baseline characteristics in control schools for grades 2 and 4, respectively. Grouped columns 3 and 5 report differences in characteristics between treatment and control schools at baseline and their standard errors, clustered by school. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

in 2006 by PASEC,<sup>9</sup> 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  $\beta_t^k$  coefficients from the following regression:

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

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  $\varepsilon_{ist}^k$  is an error term. Conditioning variables include household size, number of children, whether the head has any education, distance to school, a wealth index<sup>10</sup>, 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.<sup>11</sup>

We report standard errors, clustered at the school level, and symbols \*\*\*, \*\*, and \* to denote significance at the 1%, 5%, and 10% level of standard single hypothesis tests, respectively. In addition, since we are testing multiple hypotheses at once we compute levels of significance for each

---

<sup>9</sup>Programme for the Analysis of Education Systems of the (“Conférence des Ministres de l’Education des Etats et gouvernements de la Francophonie).

<sup>10</sup>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.

<sup>11</sup>Results without conditioning variables are presented in Appendix Table 8 and they are almost identical, but of course less precise.

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.

## 5 Results

### 5.1 Overall Treatment Effects

We begin by showing the overall effect of the program for grades 2 and 4 at baseline (they were in grade 3 and 5 at follow-up). 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 grade).

The results are in Table 2. Panel A concerns the first follow up, which was administered at the *start* of grades 3 and 5 respectively about a year after the disbursement of the project funds, while Panel B relates to the the second follow up at the *end* of grade 3 and 5 for the same children, which we are following throughout. The first three columns report effects on French, mathematics, and oral test scores. The fourth column provides a summary measure by reporting the first principal component of these three tests.<sup>12</sup> Column 5 reports PPVT scores, which were obtained only in

---

<sup>12</sup>One interpretation of the individual tests is that they are noisy measurements of one underlying human capital factor. By using the first principal component of the three tests, we may improve precision.

Table 2: Program Impacts on Grades 3 and 5 Test Scores

	<b>French</b>	<b>Math</b>	<b>Oral</b>	<b>Index</b>	<b>PPVT</b>
<b>Panel A: Beginning of Grade (First Follow-Up)</b>					
<b>Overall</b>	0.021**† (0.010)	0.019**† (0.010)	0.019*† (0.010)	0.080*† (0.044)	
Observations	5368	5361	2732	2679	
Control Mean (SD)	0.51 (0.23)	0.49 (0.23)	0.50 (0.27)		
<b>Grade 3</b>	0.029**† (0.014)	0.027**† (0.012)	0.029**† (0.014)	0.126**† (0.060)	
Observations	2720	2718	1385	1350	
Control Mean (SD)	0.53 (0.25)	0.54 (0.24)	0.35 (0.22)		
<b>Grade 5</b>	0.011 (0.011)	0.010 (0.012)	0.008 (0.013)	0.027 (0.053)	
Observations	2648	2643	1347	1329	
Control Mean (SD)	0.48 (0.20)	0.44 (0.20)	0.64 (0.24)		
<b>Panel B: End of Grade (Second Follow-Up)</b>					
<b>Overall</b>	0.020* (0.012)	0.005 (0.011)	0.026**† (0.012)	0.094*† (0.054)	0.057 (0.082)
Observations	3338	3327	1686	1620	1122
Control Mean (SD)	0.63 (0.22)	0.62 (0.22)	0.58 (0.26)		
<b>Grade 3</b>	0.035**† (0.016)	0.017 (0.015)	0.039**† (0.018)	0.160**† (0.077)	0.153 (0.096)
Observations	1732	1721	853	826	566
Control Mean (SD)	0.66 (0.23)	0.68 (0.23)	0.45 (0.23)		
<b>Grade 5</b>	0.003 (0.012)	-0.008 (0.013)	0.007 (0.014)	0.013 (0.061)	-0.060 (0.097)
Observations	1606	1606	833	794	556
Control Mean (SD)	0.59 (0.20)	0.57 (0.20)	0.72 (0.21)		

*Notes:* Standard errors are in parentheses and are adjusted for clustering. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$  correspond to p-values from the usual single-hypothesis tests. † corresponds significance at the 10% level of Romano Wolf (2005) p-values from joint tests of French, mathematics, and oral (3 tests each, by row) or to the index alone. Conditioning variables: Grade, gender, household size, number of children, education of head, distance to school, wealth index, interview language, baseline scores, missing dummies.

the second follow-up.

In the pooled sample, the index shows an improvement equal to 8.0% of a standard deviation in the first follow up, which is significant at the 5.2% level. This improvement is maintained and increased to 9.4% of a standard deviation in the second follow up, which has a p-value of 7.2%. Thus overall the program improved outcomes in the schools. When we break down the index to the individual tests we administered and adjust the p-values for multiple testing we find that all test scores improved in the first follow up by similar amounts and their adjusted p-values are less than 10%. In the second follow up the improvement in math was lost but the one in French and the oral remained and are both significant at least at the 10% level.

The improvement in overall test scores is largely driven by effects in third but not fifth grade. There are large impacts of school grants on third grade test scores across all tests. Test scores increased by almost 3 percentage points, which is a large effect in light of the means (and standard deviations) of test scores. When we look at this aggregate index of the three tests, the school grant increases third grade school performance by 0.126 of a standard deviation at the first follow-up. The effect on the index survives at 0.16 of a standard deviation through the end of grade three, indicating that the program impacts persisted two years after the grant was disbursed to schools.

It is interesting that a relatively small grant is able to improve children's learning outcomes to this extent. 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.

However, these effects are absent for fifth graders: the impacts are numerically close to zero and statistically insignificant by any criterion. The standard errors of the estimates are similar across grades, but the point estimates are much smaller.<sup>13</sup> This is perhaps surprising. However,

---

<sup>13</sup>Therefore, the lack of statistically significant results in grade 5 (but not in grade 3) does not appear to be due

principals may be investing more in earlier grades driven by a belief that learning delays emerge early in the life of the child. Indeed such a belief is actively promoted by PASEC.

Using data from the teachers' questionnaires at follow-up we investigate whether there were differential impacts of school grants on observable investments in 3<sup>rd</sup> and 5<sup>th</sup> grade students in Panel A of Appendix Table 9.<sup>14</sup> 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.

Observable parental investments are not different between grades three and five (see Panel B of Appendix Table 9), which is *prima facie* evidence that the differences are attributable to the effectiveness or administration of grants between the grades.

## 5.2 Distributional Impacts

Whether the program has different effects across the distribution is an important question relating to targeting. In Figure 3 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 (first column) and second follow-up (second column), on the treatment indicator and including the usual controls, clustered by school. The effects of the grant are generally spread over most of the distribution as shown by the index in the fourth row, although the results are less precise due to smaller sample sizes, particularly in the second follow-up. In the second follow-up, for mathematics the effects are larger at the lower end of the distribution, while for French, oral, and PPVT scores the effects are somewhat larger in the mid- to the upper end of the distribution.

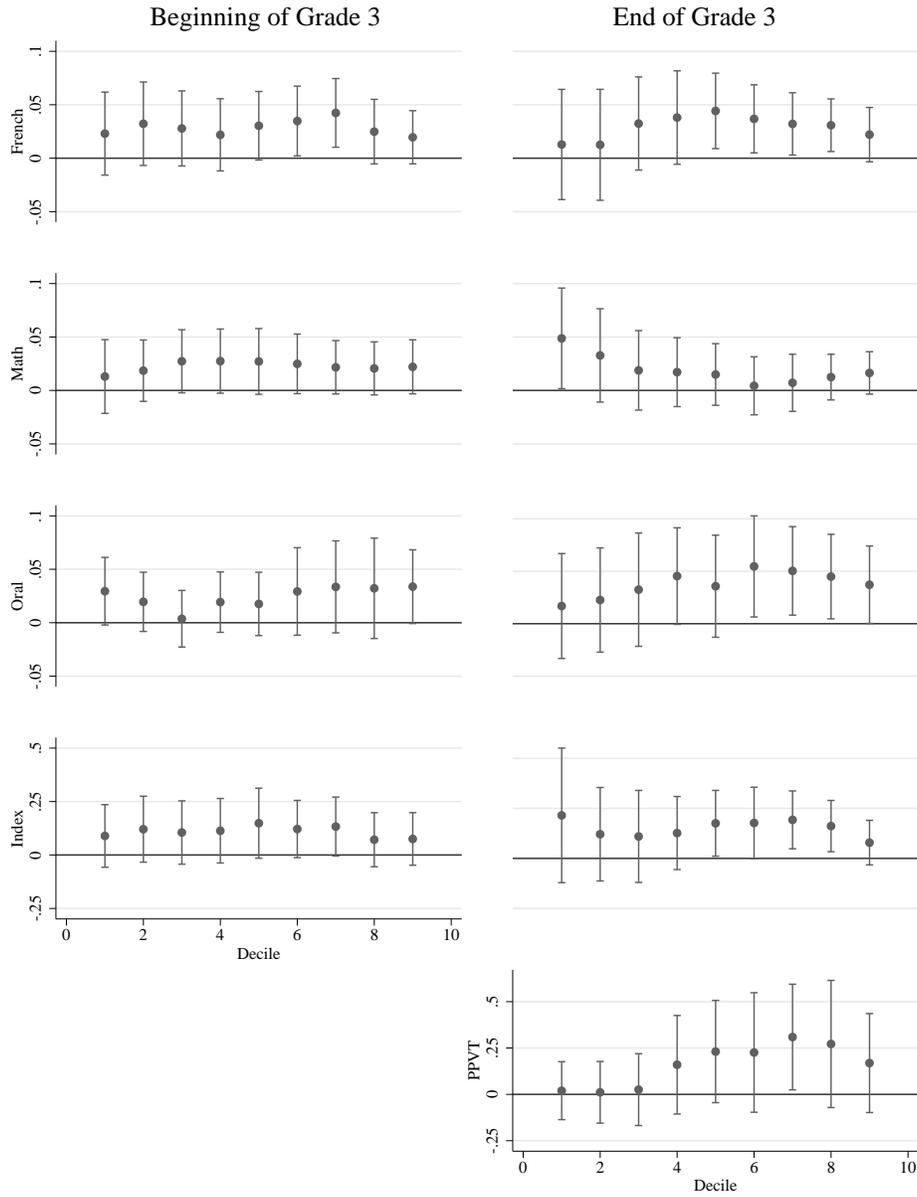
In the remaining part of the paper we look in greater detail at these results and consider

---

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.

<sup>14</sup>Ideally we would want to do this using 2<sup>nd</sup> and, say, 4<sup>th</sup> grade students, but we do not have the follow-up data for these teachers, although we have baseline data for them.

Figure 3: Distributional Impacts on Test Scores in Third Grade



Notes: Point estimates from a quantile regression at each decile with 95% confidence intervals. Index and PPVT coefficients are standardized. Beginning of grade 3 is first follow-up. End of grade three is second follow-up.

heterogeneity of effects and underlying mechanisms.

### 5.3 Heterogenous Impacts

In this section we consider characteristics by which the impact of the school grants may plausibly differ: gender, 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.<sup>15</sup> 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 related to rebel activity.

The regressions we run to construct Table 3 extend equation (1) to include an interaction between the treatment variable  $G_s$  and a pre-determined variable  $W_{ist}$  (gender, baseline ability, or region):

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

Since our larger estimates of program impacts were for students in 3<sup>rd</sup> grade, who were first exposed to the program in 2<sup>nd</sup> grade, we focus this analysis of heterogeneous impacts on them.<sup>16</sup> The results are shown in Table 3. Panel A reports results from the first follow-up, and Panel B reports results from the second follow-up. Each panel reports program impacts for each  $W_{ist}$  as well as control means and standard deviations.

There are large differences in program impact by gender. For females, the program increased test scores by 3 to 5 percentage points in the first follow-up, and increased even more in the second follow-up, with the exception of mathematics. This indicates that program impacts persisted two years after the grant was disbursed to schools for girls. 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 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). While the individual tests score differences between genders are not significant, the difference in the overall index is

---

<sup>15</sup>As mentioned, several schools were missing at baseline. In Appendix Table 10 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.

<sup>16</sup>A similar analysis performed on the test results of students in 5<sup>th</sup> grade did not produce evidence of any program impacts for this set of students (see Appendix Table 11).

Table 3: Program Impacts on Grades 3 Test Scores by Gender, Ability, and Region

	French	Math	Oral	Index	PPVT
<b>Panel A: Beginning of Grade (First Follow-Up)</b>					
<b>Male</b>	0.022 (0.017)	0.024 (0.014)	0.011 (0.017)	0.041 (0.073)	
<b>Female</b>	0.037**† (0.016)	0.031**† (0.014)	0.047***† (0.017)	0.217***† (0.073)	
Male Control Mean (SD)	0.54 (0.25)	0.56 (0.24)	0.37 (0.22)	0.03 (0.97)	
Female Control Mean (SD)	0.53 (0.24)	0.53 (0.24)	0.33 (0.22)	-0.13 (0.99)	
<b>Low Ability</b>	0.006 (0.018)	-0.007 (0.016)	0.025 (0.018)	-0.019 (0.081)	
<b>High Ability</b>	0.027 (0.020)	0.029* (0.016)	0.005 (0.019)	0.136*† (0.075)	
Low Control Mean (SD)	0.47 (0.22)	0.43 (0.20)	0.25 (0.17)	-0.46 (0.83)	
High Control Mean (SD)	0.62 (0.24)	0.68 (0.20)	0.48 (0.20)	0.49 (0.85)	
<b>North</b>	0.012 (0.016)	0.015 (0.014)	0.019 (0.016)	0.066 (0.067)	
<b>South</b>	0.102***† (0.030)	0.079***† (0.027)	0.074***† (0.028)	0.390***† (0.123)	
North Control Mean (SD)	0.56 (0.24)	0.57 (0.23)	0.38 (0.22)	0.10 (0.95)	
South Control Mean (SD)	0.41 (0.23)	0.43 (0.22)	0.23 (0.19)	-0.63 (0.88)	
<b>Panel B: End of Grade (Second Follow-Up)</b>					
<b>Male</b>	0.026 (0.018)	0.016 (0.017)	0.024 (0.022)	0.079 (0.087)	0.207* (0.115)
<b>Female</b>	0.043**† (0.019)	0.019 (0.018)	0.054**† (0.023)	0.245**† (0.102)	0.096 (0.128)
Male Control Mean (SD)	0.67 (0.23)	0.69 (0.22)	0.46 (0.22)	-0.01 (0.92)	-0.07 (0.95)
Female Control Mean (SD)	0.65 (0.24)	0.67 (0.23)	0.43 (0.23)	-0.17 (1.03)	-0.10 (1.01)
<b>Low Ability</b>	0.027 (0.022)	0.001 (0.022)	0.036 (0.026)	0.081 (0.116)	
<b>High Ability</b>	0.029 (0.022)	0.027 (0.019)	0.028 (0.024)	0.174*† (0.096)	
Low Control Mean (SD)	0.60 (0.23)	0.60 (0.22)	0.34 (0.20)	-0.45 (0.93)	
High Control Mean (SD)	0.75 (0.20)	0.78 (0.19)	0.56 (0.20)	0.39 (0.82)	
<b>North</b>	0.024 (0.018)	0.000 (0.016)	0.023 (0.020)	0.087 (0.084)	0.130 (0.097)
<b>South</b>	0.079**† (0.037)	0.084**† (0.033)	0.105***† (0.035)	0.450***† (0.161)	0.181 (0.250)
North Control Mean (SD)	0.69 (0.23)	0.71 (0.22)	0.48 (0.22)	0.07 (0.92)	-0.21 (0.88)
South Control Mean (SD)	0.58 (0.24)	0.57 (0.23)	0.31 (0.21)	-0.67 (0.96)	0.43 (1.13)

*Notes:* Standard errors are in parentheses and are adjusted for clustering. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$  correspond to p-values from the usual single-hypothesis tests. † corresponds significance at the 10% level of Romano Wolf (2005) p-values from joint tests of French, mathematics, and oral (3 tests each, by row) or to the index alone. Conditioning variables: Grade, gender, household size, number of children, education of head, distance to school, wealth index, interview language, baseline scores, missing dummies.

significant at least at the 10% level.

There are also several education interventions that benefit mostly girls. It is much less common to find programs that affect boys alone. Some examples of (early childhood) interventions in developed countries that produce larger cognitive and schooling effects in girls than boys are reviewed in Anderson (2008) (see also the results in Heckman et al. (2010), or Ramey and Campbell (2007), regarding education outcomes of these interventions). Similarly, Krueger (1999) also reports that the STAR class size experiment produce smaller short run impacts, but larger cumulative impacts for girls than boys, and Chetty et al. (2014) show slightly larger long term impacts of teacher quality on girls than boys. Although this was not directly an educational intervention (but which may have partly operated through access to better schools), the Moving to Opportunity experiment studied in Kling et al. (2007) also shows much stronger impacts for girls. In developing countries, several papers show stronger impacts of interventions on girls than boys (Kremer and Holla (2009)), although these concern primarily interventions that increase access to school.

At baseline girls score between 10% - 20% of a standard deviation below boys in the cognitive tests we administer (not shown) and hence start from a lower base. However, this explanation of larger program impacts is hard to reconcile with the fact that, as we show below, effects are larger for those with higher baseline ability (and this is especially true for females). An alternative hypothesis would be that girls bring to elementary schools more discipline, patience, and higher levels of maturity overall than boys at a given age, 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.

The program also had a large impact for higher-ability students: the index of scores (column 4) increased by 0.14 standard deviations in the first follow-up and 0.17 standard deviations in the second follow-up as a result of the program, though the coefficients are only marginally significant. 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. However, despite the large changes in the point estimates across ability groups, the differences are not significant.

We now turn to differences by North and South of the country, two very different regions. 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.<sup>17</sup> In fact, if we focus on 3<sup>rd</sup> 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. For example, as a result of the program, students in southern schools are able to increase the proportion of correct answers by 10.2 percentage points, which is almost 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.

The South-North differences in estimates of the impact of school grants are striking as well as highly significant overall for the first follow up (p-value 2.1% for the index). 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.

#### **5.4 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 4, Panel A. Students in the southern schools perform worse on almost all tests than their counterparts in the North. For control schools in the first follow-up, documented in Panel B, the differences between the North and the South are even larger.

As mentioned above, at baseline we were only able to survey a subsample of schools. The missing schools (recovered at follow-up) were, as far as we can see, balanced in their treatment and control status, but they were different from the sampled schools. In fact, as we report in the appendix (Appendix Table 13), among control schools, missing schools are worse than the non-missing schools on a number of time-invariant dimensions, as one might expect.

Therefore, it is probably safe to say that, once we look at the schools in the follow-up which we are using to measure program impacts, the schools in the South show much lower test results than the schools in the North.

---

<sup>17</sup>In Appendix Table 12 we show that the samples are balanced within each geographic region.

Table 4: Regional Differences, Second-Third Grade

	South	North	Difference	
<b>Panel A: Test Scores at Beginning of Second Grade (Baseline)</b>				
Percent Correct: French	0.430	0.420	0.010	(0.027)
Percent Correct: Math	0.325	0.373	-0.048**	(0.023)
Percent Correct: Oral	0.154	0.242	-0.088***	(0.016)
Index Score (standardized)	-0.239	0.049	-0.288**	(0.112)
<b>Panel B: Test Scores at Beginning of Third Grade (First Follow-Up, Control Schools)</b>				
Percent Correct: French	0.411	0.564	-0.153***	(0.022)
Percent Correct: Math	0.434	0.569	-0.136***	(0.021)
Percent Correct: Oral	0.233	0.383	-0.150***	(0.022)
Index Score (standardized)	-0.629	0.100	-0.730***	(0.100)
<b>Panel C: Household Characteristics (First Follow-Up, Control Schools)</b>				
Household size	8.625	10.216	-1.591***	(0.412)
Number of children in household	5.050	5.551	-0.501*	(0.276)
Head has any education	0.550	0.401	0.149***	(0.050)
Percent of adult females with any education	0.261	0.224	0.038	(0.043)
Wealth index	-0.654	0.137	-0.792***	(0.092)
Interview conducted in French	0.175	0.090	0.085**	(0.041)
<b>Panel D: Project Characteristics (Second Follow-Up, Treatment Schools)</b>				
Months since project began	15.914	23.479	-7.564***	(1.144)
Students helped draft application	0.800	0.547	0.253***	(0.082)
Project included manuals	0.800	0.895	-0.095	(0.074)
Project included computer materials	0.029	0.121	-0.092**	(0.042)
Project included teacher training	0.914	0.752	0.162**	(0.062)
Project included management training	0.629	0.368	0.261***	(0.093)
Project included building courses	0.971	0.821	0.151***	(0.046)
Project included improving general education	0.563	0.456	0.106	(0.100)
Project included improving educational outputs	0.114	0.129	-0.015	(0.063)
Amount spent on principal (1,000,000s CFA)	0.082	0.034	0.048***	(0.014)
Amount spent on teachers (1,000,000s CFA)	0.317	0.278	0.039	(0.058)
Amount spent on management (1,000,000s CFA)	0.128	0.041	0.087***	(0.022)
Amount spent on students (1,000,000s CFA)	0.505	1.025	-0.520***	(0.092)

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

Panel C 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).

Finally, Panel D 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 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 in particular computer related materials, while schools in the South were much more likely to have components related to training of teachers, building courses, managerial training, spending on the principal and the teachers. At the same time the Northern schools reported more spending on 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 equipment. This may well be a force behind the large differences in program impacts in these two sets of schools.

Table 5 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).

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 an explicit pedagogical emphasis, which did not (in the government's definition) include physical infrastructure. 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 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 4. 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, 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.

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.

Panel B shows program impacts on teacher and classroom characteristics as reported by the 3rd grade teacher in the first follow-up. The number of manuals are not reported by the teacher as having increased significantly either in the north or the south, despite the impact on manuals reported above, and measuring equipment in the South is reported to have increased as a response to the program, but not in the North.

Table 5: Program Impacts on School Characteristics by Region, First Follow-Up

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

Finally, the behavior of students is reported to have improved considerably in the South, but not in the North: treatment affected how often teachers had to ask for silence during the day in southern schools. This corresponds to what we found before when we compared the reports of 3<sup>rd</sup> and 5<sup>th</sup> grade teacher (see Appendix Table 9). Student behavior improved among 3<sup>rd</sup> graders but not among 5<sup>th</sup> graders, which is exactly what happened in terms of test results. The improved behavior, may be an outcome driven by improved teaching more generally, which underlies the improved scores.

We also examined the impact of the program on household behaviors in the South and in the North, which is shown in Appendix Table 13. However, 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 in the south, and projects in the south tended to focus on training and human resources 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, although much of the research concerns developed countries and the US in particular. 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 and indeed they are encouraged to do so. 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 information technology (IT) and other educational 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)).

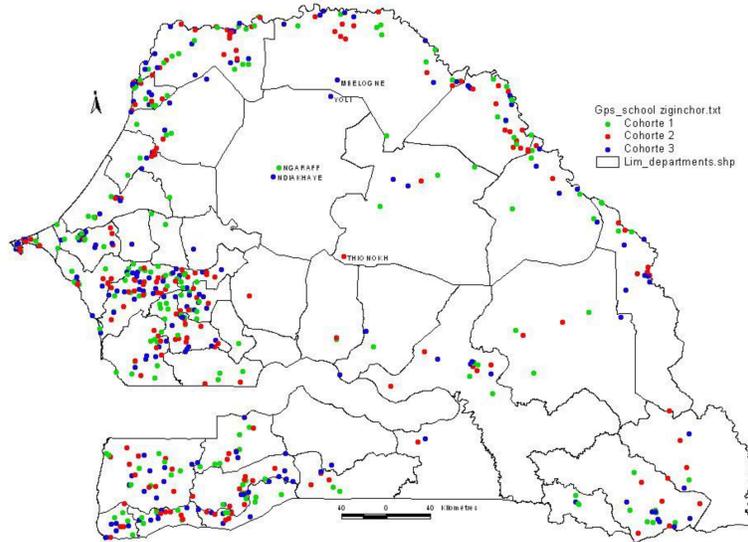
## References

- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484). 5.3
- ANSD (2007). La pauvreté au Sénégal: de la dévaluation de 1994 à 2001-2002. Technical report, Agence Nationale de la Statistique et de la Démographie. 5.3
- 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. 1
- Blimpo, M., Evans, D., and Lahire, N. (2014). School-based management and educational outcomes, lessons from a randomized field experiment. 1
- Bruns, B., Filmer, D., and Patrinos, H. A. (2011). *Making schools work: New evidence on accountability reforms*. World Bank Publications. 5.1
- Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, 104(9):2633–2679. 5.3
- 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. 1
- Galiani, S. and Perez-Truglia, R. (2013). School management in developing countries. *Education Policy in Developing Countries*, 4:193. 1
- 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. 1
- Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2:945–1017. 1
- 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. 1
- 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. 1
- Hanushek, E. A. (2006). School resources. *Handbook of the Economics of Education*, 2:865–908. 5.1
- 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. 1
- Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, 2:1051–1078. 6

- Hanushek, E. A. and Woessmann, L. (2010). The economics of international differences in educational achievement. Technical report, National Bureau of Economic Research. 1
- Heckman, J., Moon, S., Pinto, R., Savelyev, P., and Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the highscope perry preschool program. *Quantitative economics*, 1(1):1–46. 5.3
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119. 5.3
- Kremer, M. and Holla, A. (2009). Improving education in the developing world: what have we learned from randomized evaluations? *Annual Review of Economics*, 1:513. 5.3
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, pages 497–532. 5.3
- 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. 1
- PASEC (2007). Evaluation pascal senegal. Technical report, Program on the Analysis of Education Systems. 1
- Ramey, C. T. and Campbell, F. A. (2007). Carolina abecedarian project. Technical report, <http://phoenixday.org/wp-content/uploads/2012/04/res-Abecedarian-studies-full.pdf>. 5.3
- 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. 5.1
- Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282. 1
- World Bank (2013). Projet d’amélioration de la qualité et de l’équité de l’éducation de base. Technical report. 1

# Online Appendix: Not for Publication

Figure 4: Location of Schools in Sample

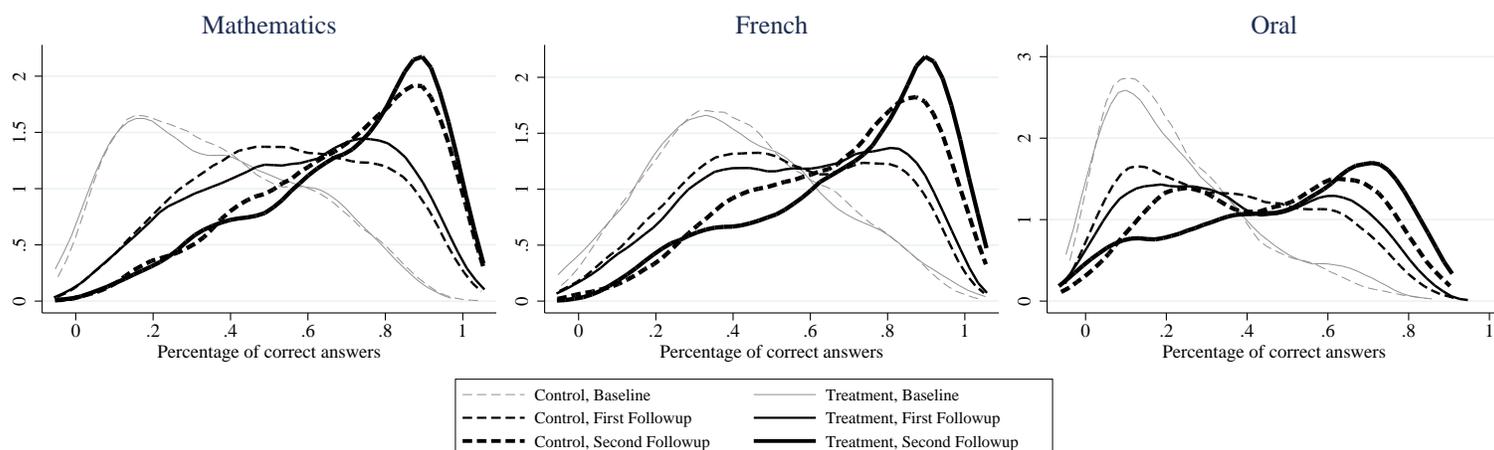


Note: The allocation of schools to cohorts is the result of the randomization of the sequence with which grants were allocated. Cohort 1 were the first to receive grants, followed by cohort 2 and then cohort 3.

## A. Test Score Distributions

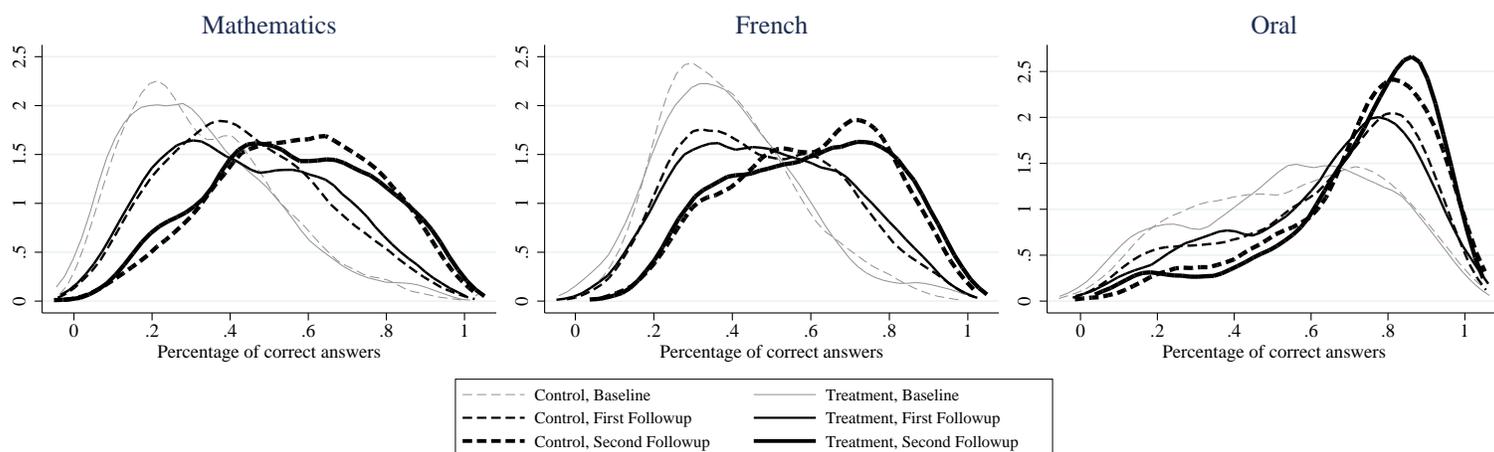
In this appendix we document the way the distribution of test scores changed over time and as a result of the experiment. Figure 5 shows the densities of scores in the French, mathematics, and oral tests taken at the beginning of 2nd grade at baseline, the beginning of 3rd grade at first follow-up, and the end of 3rd grade at second follow-up. The dashed lines in the figures correspond to the control schools while the solid lines are the treatment schools.

Figure 5: Distribution of Second/Third Grade Scores



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). The lines with a medium level of thickness correspond to the densities of scores at the beginning of third grade at first follow-up. Although the densities of test scores are balanced at baseline, they are different at first follow-up, 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. The thickest lines in Figure 5 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 towards the end of the third grade, roughly two years after the funds were disbursed to the treatment schools.

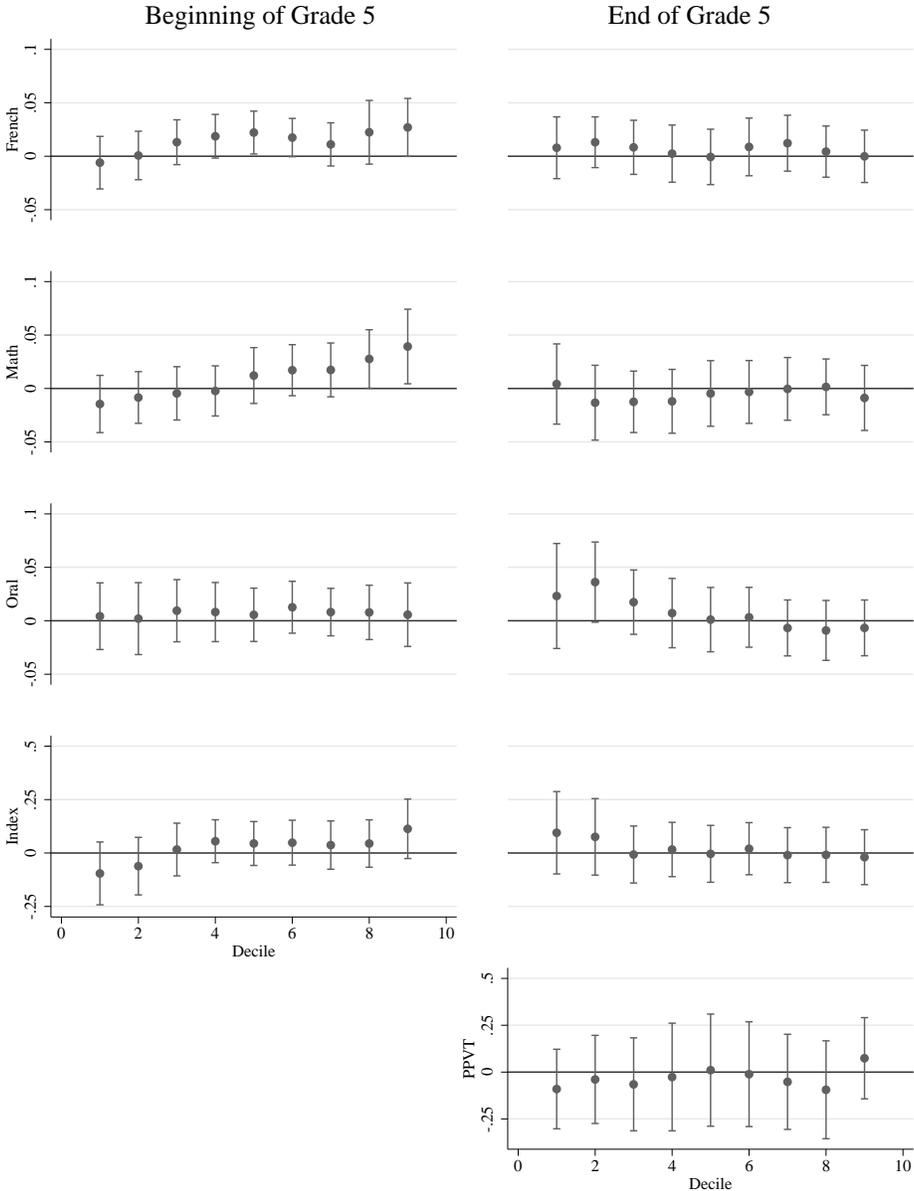
Figure 6: Distribution of Fourth/Fifth Grade Scores



Analogously, in Figure 6, we display the densities of scores at first and second follow-up for students 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. In the main body of the paper we show that there is no statistically significant impact of school grants on fifth grade test scores.

B. Other Appendix Tables and Figures

Figure 7: Distributional Impacts on Test Scores in Fifth Grade



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

Table 6: First Follow-up Descriptive Statistics and Balance

	Grade 2				Grade 4			
	Control		Treat-Control		Control		Treat-Control	
<b>Panel A: Household Characteristics</b>								
Household size	9.92	(4.41)	0.16	(0.34)	10.04	(4.31)	-0.46	(0.31)
Number of children in household	5.46	(2.80)	0.24	(0.21)	5.69	(2.79)	-0.32	(0.21)
Head has any education	0.43	(0.50)	-0.00	(0.04)	0.44	(0.50)	-0.03	(0.04)
Percent adult females with any education	0.23	(0.34)	0.00	(0.03)	0.24	(0.36)	0.02	(0.03)
Distance to school (km)	0.61	(0.63)	-0.09**	(0.04)	0.59	(0.60)	0.01	(0.04)
Home has electricity	0.44	(0.50)	0.05	(0.04)	0.44	(0.50)	0.02	(0.04)
Home has modern toilet	0.37	(0.48)	-0.02	(0.04)	0.35	(0.48)	-0.03	(0.04)
Land owned (hectares)	2.59	(4.63)	0.53	(0.48)	2.33	(3.49)	0.09	(0.29)
Interview conducted in French	0.11	(0.31)	0.01	(0.02)	0.09	(0.29)	-0.00	(0.02)
<b>Panel B: School Characteristics</b>								
Distance to nearest city (km)	18.35	(24.78)	0.10	(2.10)	17.41	(22.61)	0.91	(2.07)
Population in locality (100,000s)	0.92	(2.74)	0.27	(0.31)	0.93	(2.75)	0.30	(0.32)
Locality has health center	0.71	(0.45)	0.03	(0.04)	0.71	(0.45)	0.03	(0.04)
School located in South	0.19	(0.40)	0.00	(0.04)	0.20	(0.40)	-0.00	(0.04)

*Notes:* Grouped columns 2 and 4 report means and standard deviations of baseline characteristics in control schools for grades 2 and 4, respectively. Grouped columns 3 and 5 report differences in characteristics between treatment and control schools at baseline and their standard errors, clustered by school. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 7: Summary Statistics of Nationally Representative Sample of 2nd and 5th Grade Students

	Mean	Standard Deviation
<b>Panel A: School and Teacher Characteristics</b>		
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)
<b>Panel B: Household Characteristics</b>		
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)

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

Table 8: Program Impacts on Grades 3 and 5 Test Scores, No Controls

	<b>French</b>	<b>Math</b>	<b>Oral</b>	<b>Index</b>	<b>PPVT</b>
<b>Panel A: Beginning of Grade (First Follow-Up)</b>					
<b>Overall</b>	0.021 (0.014)	0.018 (0.014)	0.017 (0.015)	0.078 (0.068)	
Observations	5368	5361	2732	2679	
Control Mean (SD)	0.51 (0.23)	0.49 (0.23)	0.50 (0.27)		
<b>Grade 3</b>	0.029* (0.017)	0.026 (0.016)	0.032* (0.018)	0.138* (0.081)	
Observations	2720	2718	1385	1350	
Control Mean (SD)	0.53 (0.25)	0.54 (0.24)	0.35 (0.22)		
<b>Grade 5</b>	0.011 (0.015)	0.009 (0.016)	0.004 (0.018)	0.016 (0.081)	
Observations	2648	2643	1347	1329	
Control Mean (SD)	0.48 (0.20)	0.44 (0.20)	0.64 (0.24)		
<b>Panel B: End of Grade (Second Follow-Up)</b>					
<b>Overall</b>	0.019 (0.015)	0.003 (0.015)	0.026 (0.017)	0.091 (0.077)	0.046 (0.085)
Observations	3338	3327	1686	1620	1122
Control Mean (SD)	0.63 (0.22)	0.62 (0.22)	0.58 (0.26)		
<b>Grade 3</b>	0.033* (0.019)	0.014 (0.018)	0.045** (0.022)	0.176* (0.096)	0.170* (0.098)
Observations	1732	1721	853	826	566
Control Mean (SD)	0.66 (0.23)	0.68 (0.23)	0.45 (0.23)		
<b>Grade 5</b>	0.002 (0.016)	-0.011 (0.018)	0.007 (0.018)	0.002 (0.093)	-0.082 (0.097)
Observations	1606	1606	833	794	556
Control Mean (SD)	0.59 (0.20)	0.57 (0.20)	0.72 (0.21)		

Notes: Standard errors are in parentheses and are adjusted for clustering. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$  correspond to p-values from the usual single-hypothesis tests.

Table 9: Program Impacts on Teacher and Household Outcomes, First Follow-Up

	<b>Grade 3</b>	<b>Grade 5</b>	<b>Difference</b>
<b>Panel A: Teacher Outcomes</b>			
Teacher has Baccalaureate	-0.052 (0.047)	0.018 (0.046)	-0.070 (0.064)
Teacher had training in past 5 years	0.083** (0.039)	0.101** (0.047)	-0.018 (0.056)
Minutes spent preparing lesson	2.115 (1.726)	0.061 (1.614)	2.054 (2.023)
Number of manuals	6.209 (4.231)	6.493 (4.964)	-0.284 (5.296)
Number of measuring instruments	0.138 (0.179)	0.228 (0.193)	-0.090 (0.210)
Number of chairs	0.018 (0.038)	0.071** (0.035)	-0.053 (0.041)
Teaches with books	0.004 (0.013)	0.012 (0.012)	-0.008 (0.015)
Teaches with computers	-0.005 (0.019)	0.034 (0.024)	-0.038 (0.029)
Times per day ask for silence	-1.638** (0.686)	1.109 (0.830)	-2.747*** (0.931)
Times per day punish a student	-0.126 (0.277)	0.741** (0.370)	-0.868** (0.407)
Number of students who left in past year	0.303 (0.730)	1.063 (0.800)	-0.760 (1.095)
Number of students who joined in past year	0.262 (0.212)	-0.030 (0.185)	0.292 (0.194)
<b>Panel B: Household Outcomes</b>			
Days of school missed last week	0.106 (0.075)	0.010 (0.055)	0.096 (0.086)
Student works after school	-0.008 (0.012)	-0.016 (0.014)	0.008 (0.014)
Parent involved in school	0.037 (0.038)	0.026 (0.038)	0.011 (0.039)
Expenditure on uniform (1,000s CFA)	-0.003 (0.070)	-0.070 (0.056)	0.067 (0.055)
Expenditure on tuition (1,000s CFA)	0.303 (0.303)	0.035 (0.164)	0.269 (0.313)
Expenditure on supplies (1,000s CFA)	-0.162 (0.232)	-0.111 (0.320)	-0.050 (0.302)
Student has tutor	-0.027 (0.026)	-0.009 (0.029)	-0.017 (0.031)
Expenditure on children (1,000s CFA)	0.303 (0.611)	-0.175 (0.662)	0.477 (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 10: Characteristics of Schools by Baseline Missing Status, First Follow-Up

	Not Missing	Missing	Difference
<b>Panel A: Treatment Status</b>			
Treated	0.337 (0.473)	0.312 (0.464)	0.025 (0.089)
<b>Panel B: Control School Characteristics</b>			
School located in South	0.186 (0.389)	0.439 (0.497)	-0.253** (0.116)
School located in Rural Area	0.736 (0.441)	1.000 (0.000)	-0.264*** (0.025)
Locality Population (100,000s)	0.981 (2.827)	0.024 (0.026)	0.957*** (0.173)
Number of Teachers	9.946 (4.954)	7.090 (4.144)	2.856*** (0.991)
Number of Pupils	341.617 (254.680)	238.326 (152.758)	103.291*** (37.826)
Percent Correct: French, Grade 3	0.540 (0.245)	0.418 (0.239)	0.123*** (0.045)
Percent Correct: Math, Grade 3	0.550 (0.236)	0.408 (0.228)	0.142*** (0.042)
Percent Correct: Oral, Grade 3	0.361 (0.221)	0.212 (0.178)	0.149*** (0.037)
Percent Correct: Index, Grade 3	-0.012 (1.009)	-0.666 (0.939)	0.653*** (0.203)
Percent Correct: French, Grade 5	0.483 (0.201)	0.379 (0.190)	0.104*** (0.032)
Percent Correct: Math, Grade 5	0.441 (0.204)	0.342 (0.195)	0.099*** (0.037)
Percent Correct: Oral, Grade 5	0.648 (0.236)	0.572 (0.257)	0.076 (0.050)
Percent Correct: Index, Grade 5	0.026 (0.943)	-0.483 (0.937)	0.509*** (0.179)

Columns 1 and 2 report means and standard deviations (in parentheses) for schools that were not missing at baseline and missing at baseline, respectively. Column 3 reports the difference in means between not missing and missing schools, and clustered standard errors (at the school level) in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 11: Program Impacts on Grades 5 Test Scores by Gender, Ability, and Region

	French	Math	Oral	Index	PPVT
<b>Panel A: Beginning of Grade (First Follow-Up)</b>					
<b>Male</b>	0.014 (0.012)	0.008 (0.013)	0.020 (0.015)	0.043 (0.064)	
<b>Female</b>	0.009 (0.013)	0.013 (0.013)	-0.005 (0.015)	0.009 (0.061)	
Male Control Mean (SD)	0.49 (0.20)	0.46 (0.21)	0.66 (0.24)	0.07 (1.02)	
Female Control Mean (SD)	0.46 (0.20)	0.41 (0.20)	0.63 (0.23)	-0.09 (0.96)	
<b>Low Ability</b>	0.009 (0.012)	-0.009 (0.013)	0.003 (0.019)	-0.016 (0.066)	
<b>High Ability</b>	0.000 (0.016)	0.013 (0.018)	0.017 (0.013)	0.039 (0.071)	
Low Control Mean (SD)	0.35 (0.14)	0.33 (0.15)	0.51 (0.22)	-0.63 (0.76)	
High Control Mean (SD)	0.62 (0.16)	0.56 (0.18)	0.79 (0.13)	0.72 (0.66)	
<b>North</b>	0.017 (0.011)	0.014 (0.013)	-0.002 (0.013)	0.029 (0.057)	
<b>South</b>	-0.021 (0.026)	-0.010 (0.030)	0.049 (0.032)	0.009 (0.130)	
North Control Mean (SD)	0.50 (0.20)	0.46 (0.20)	0.68 (0.22)	0.13 (0.96)	
South Control Mean (SD)	0.37 (0.17)	0.33 (0.18)	0.52 (0.25)	-0.55 (0.94)	
<b>Panel B: End of Grade (Second Follow-Up)</b>					
<b>Male</b>	-0.001 (0.014)	-0.013 (0.015)	0.019 (0.018)	0.046 (0.081)	-0.029 (0.123)
<b>Female</b>	0.008 (0.013)	-0.003 (0.014)	-0.006 (0.018)	-0.022 (0.072)	-0.095 (0.127)
Male Control Mean (SD)	0.60 (0.20)	0.59 (0.20)	0.73 (0.20)	0.06 (0.99)	0.05 (1.02)
Female Control Mean (SD)	0.59 (0.19)	0.54 (0.21)	0.71 (0.21)	-0.06 (0.95)	0.03 (1.01)
<b>Low Ability</b>	-0.002 (0.015)	-0.020 (0.016)	-0.032 (0.023)	-0.120 (0.087)	
<b>High Ability</b>	0.006 (0.014)	-0.007 (0.018)	0.012 (0.015)	0.016 (0.079)	
Low Control Mean (SD)	0.49 (0.17)	0.46 (0.18)	0.63 (0.21)	-0.53 (0.86)	
High Control Mean (SD)	0.71 (0.15)	0.68 (0.17)	0.83 (0.12)	0.62 (0.65)	
<b>North</b>	0.003 (0.013)	-0.009 (0.014)	0.004 (0.015)	0.004 (0.067)	-0.014 (0.099)
<b>South</b>	-0.002 (0.026)	-0.010 (0.028)	0.020 (0.036)	0.046 (0.144)	-0.181 (0.259)
North Control Mean (SD)	0.62 (0.19)	0.60 (0.19)	0.74 (0.19)	0.11 (0.92)	-0.06 (0.94)
South Control Mean (SD)	0.49 (0.20)	0.45 (0.20)	0.65 (0.24)	-0.47 (1.04)	0.44 (1.18)

*Notes:* Standard errors are in parentheses and are adjusted for clustering. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$  correspond to p-values from the usual single-hypothesis tests. † corresponds significance at the 10% level of Romano Wolf (2005) p-values from joint tests of French, mathematics, and oral (3 tests each, by row) or to the index alone. Conditioning variables: Grade, gender, household size, number of children, education of head, distance to school, wealth index, interview language, baseline scores, missing dummies.

Table 12: Baseline Descriptive Statistics and Balance (Grade 2), by Region

	North				South			
	Control		Treat-Control		Control		Treat-Control	
<b>Panel A: Test Scores</b>								
Percent Correct: French	0.42	(0.21)	0.01	(0.02)	0.46	(0.26)	-0.08	(0.05)
Percent Correct: Math	0.37	(0.23)	0.00	(0.02)	0.33	(0.23)	-0.02	(0.05)
Percent Correct: Oral	0.24	(0.17)	0.01	(0.02)	0.14	(0.14)	0.03	(0.03)
Index Score (standardized)	0.04	(0.98)	0.04	(0.09)	-0.17	(0.96)	-0.19	(0.23)
<b>Panel B: Household Characteristics</b>								
Days of school missed last week	0.18	(0.91)	0.06	(0.08)	0.10	(0.55)	0.10	(0.13)
Student works after school	0.01	(0.11)	-0.00	(0.01)	0.00	(0.00)	0.04	(0.02)
Household size	9.14	(4.03)	0.11	(0.36)	9.79	(4.14)	-0.49	(0.69)
Number of children in household	5.13	(2.56)	0.06	(0.23)	5.80	(2.76)	-0.46	(0.47)
Head has any education	0.58	(0.49)	0.00	(0.04)	0.68	(0.47)	-0.10	(0.09)
Percent of adult females with any education	0.37	(0.41)	-0.03	(0.03)	0.35	(0.39)	-0.01	(0.07)
Distance to school (km)	0.68	(0.86)	-0.00	(0.07)	0.88	(1.06)	-0.38***	(0.12)
Parent involved in school	0.41	(0.49)	0.07	(0.04)	0.25	(0.43)	0.17*	(0.08)
Expenditure on household food (1,000s CFA)	23.56	(16.04)	1.19	(1.29)	14.13	(9.22)	1.25	(1.87)
Expenditure on uniform (1,000s CFA)	2.26	(1.10)	0.10	(0.41)	3.75	(1.04)	-0.72	(0.70)
Expenditure on tuition (1,000s CFA)	1.08	(1.17)	-0.00	(0.10)	1.19	(1.22)	-0.01	(0.23)
Expenditure on supplies (1,000s CFA)	4.10	(6.14)	-0.46	(0.34)	2.71	(1.95)	-0.05	(0.35)
Student has tutor	0.18	(0.38)	-0.01	(0.03)	0.04	(0.18)	0.00	(0.03)
Home has electricity	0.52	(0.50)	0.02	(0.05)	0.25	(0.43)	0.08	(0.09)
Home has modern toilet	0.60	(0.49)	-0.01	(0.04)	0.29	(0.46)	-0.04	(0.08)
Land owned (hectares)	2.42	(3.53)	0.27	(0.52)	2.15	(3.18)	1.15	(0.99)
Interview conducted in French	0.10	(0.30)	-0.03	(0.02)	0.22	(0.42)	-0.08	(0.06)
<b>Panel C: School and Teacher Characteristics</b>								
Distance to nearest city (km)	18.83	(26.53)	0.08	(2.56)	16.38	(16.54)	-0.87	(3.10)
Locality population (100,000s)	1.67	(4.83)	0.04	(0.56)	0.14	(0.37)	-0.04	(0.07)
Locality has health center	0.70	(0.46)	0.03	(0.05)	0.74	(0.44)	-0.01	(0.10)
School has Electricity	0.58	(0.49)	0.02	(0.05)	0.50	(0.50)	-0.01	(0.11)
Number of Teachers	9.53	(4.88)	0.61	(0.56)	10.35	(5.29)	-0.30	(1.23)
Number of Pupils	335.33	(252.92)	36.62	(28.67)	366.65	(248.77)	-8.09	(56.47)
School has library	0.24	(0.43)	0.07	(0.05)	0.07	(0.25)	0.10	(0.08)
Number of computers	1.44	(4.48)	-0.03	(0.46)	0.58	(3.87)	0.06	(0.66)
Number of manuals in classroom	61.41	(45.68)	5.07	(5.15)	53.02	(42.26)	-5.60	(9.26)
Percent teachers female	0.33	(0.23)	0.01	(0.03)	0.24	(0.22)	-0.01	(0.04)
Average teacher age	33.16	(4.45)	-0.16	(0.44)	32.93	(3.18)	-0.00	(0.77)
Percent teachers with Bacculaureate	0.41	(0.23)	-0.01	(0.03)	0.42	(0.21)	-0.05	(0.05)
Average teacher experience	6.57	(3.90)	0.17	(0.41)	6.49	(2.57)	-0.34	(0.57)
Percent teachers with training in past 5 years	0.50	(0.50)	0.10*	(0.05)	0.37	(0.48)	0.09	(0.12)
Percent of principals with Bacculaureate	0.74	(0.44)	-0.02	(0.05)	0.73	(0.44)	-0.19*	(0.11)

Notes: Grouped columns 2 and 4 report means and standard deviations of baseline characteristics in control schools for grades 2 and 4, respectively. Grouped columns 3 and 5 report differences in characteristics between treatment and control schools at baseline and their standard errors, clustered by school. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 13: Program Impacts on Grade 3 Household Characteristics by Region, First Follow-Up

	<b>South</b>	<b>North</b>	<b>Difference</b>
Student works after school	0.015 (0.025)	-0.013 (0.014)	0.017 (0.025)
Days of school missed last week	0.345** (0.173)	0.049 (0.082)	0.391 (0.291)
Parent involved in school	-0.039 (0.073)	0.057 (0.043)	-0.017 (0.117)
Expenditure on uniform (1,000s CFA)	0.008 (0.257)	-0.010 (0.059)	-0.015 (0.254)
Expenditure on tuition (1,000s CFA)	1.321 (1.431)	0.056 (0.117)	1.230 (1.427)
Expenditure on supplies (1,000s CFA)	-0.129 (0.476)	-0.155 (0.261)	-0.015 (0.640)
Student has tutor	-0.018 (0.029)	-0.027 (0.031)	0.014 (0.048)

*Notes:* Columns 1 and 2 report program impacts in the South and North, and Column 3 reports the difference in program impacts between South and North. Clustered standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 14: 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

*Notes:* 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).

*Notes:* 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 15: Difference in Baseline Characteristics among Non-Attriters

	Grade 2				Grade 4			
	1st Followup		2nd Followup		1st Followup		2nd Followup	
<b>Panel A: Test Scores</b>								
Percent Correct: French	-0.005	(0.018)	0.006	(0.022)	0.000	(0.014)	0.001	(0.016)
Percent Correct: Math	-0.005	(0.018)	-0.000	(0.021)	-0.003	(0.016)	-0.004	(0.018)
Percent Correct: Oral	0.012	(0.015)	0.009	(0.018)	-0.007	(0.019)	-0.006	(0.023)
Index Score (standardized)	-0.020	(0.095)	0.052	(0.114)	-0.046	(0.083)	-0.029	(0.098)
<b>Panel B: Household Characteristics</b>								
Distance to nearest city (km)	0.347	(2.203)	-0.513	(2.436)	0.035	(2.341)	-0.985	(2.579)
Locality population (100,000s)	-0.063	(0.407)	0.160	(0.442)	0.142	(0.473)	0.295	(0.500)
Locality has health center	0.032	(0.044)	0.076	(0.052)	0.032	(0.043)	0.069	(0.052)
School located in South	-0.016	(0.038)	-0.024	(0.046)	-0.021	(0.037)	-0.040	(0.044)
School has Electricity	0.021	(0.049)	-0.019	(0.056)	0.037	(0.049)	0.003	(0.057)
Number of Teachers	0.464	(0.515)	0.636	(0.587)	0.534	(0.515)	0.708	(0.609)
Number of Pupils	33.657	(25.444)	43.355	(28.603)	31.181	(26.044)	47.491	(30.330)
School has library	0.085**	(0.043)	0.087*	(0.047)	0.094**	(0.044)	0.079	(0.049)
Number of computers	-0.000	(0.397)	-0.144	(0.485)	0.052	(0.409)	-0.170	(0.515)
Number of manuals in classroom	3.810	(4.564)	3.572	(5.758)	6.530	(5.498)	10.830*	(6.374)
Percent teachers female	0.007	(0.023)	0.028	(0.026)	0.012	(0.023)	0.032	(0.026)
Average teacher age	-0.126	(0.394)	-0.056	(0.459)	0.051	(0.392)	0.020	(0.454)
Percent teachers with Baccaalaureate	-0.012	(0.023)	-0.016	(0.027)	-0.017	(0.023)	-0.017	(0.027)
Average teacher experience	0.085	(0.345)	0.138	(0.394)	0.220	(0.355)	0.168	(0.404)
Percent teachers with training in past 5 years	0.092*	(0.050)	0.071	(0.058)	0.012	(0.050)	0.024	(0.058)
Percent principals with Baccaalaureate	-0.034	(0.045)	-0.022	(0.052)	-0.037	(0.045)	-0.026	(0.053)
<b>Panel C: School and Teacher Characteristics</b>								
Days of school missed last week	0.084	(0.077)	0.068	(0.109)	-0.078*	(0.042)	-0.139**	(0.058)
Student works after school	0.009	(0.009)	0.009	(0.011)	-0.015*	(0.008)	-0.025*	(0.013)
Household size	-0.126	(0.335)	0.498	(0.373)	0.217	(0.356)	0.111	(0.430)
Number of children in household	-0.098	(0.214)	0.125	(0.242)	0.249	(0.239)	0.159	(0.289)
Head has any education	-0.030	(0.039)	-0.023	(0.047)	-0.077*	(0.040)	-0.067	(0.048)
Percent of adult females with any education	-0.016	(0.034)	-0.010	(0.040)	-0.010	(0.030)	0.002	(0.036)
Distance to school (km)	-0.093	(0.064)	-0.125	(0.076)	-0.020	(0.134)	-0.143	(0.239)
Parent involved in school	0.070*	(0.040)	0.081*	(0.048)	-0.069*	(0.040)	-0.066	(0.049)
Expenditure on household food (1,000s CFA)	0.948	(1.236)	2.343	(1.449)	1.216	(1.286)	0.372	(1.575)
Expenditure on uniform (1,000s CFA)	0.226	(0.437)	0.033	(0.593)	-0.124	(0.336)	-0.234	(0.410)
Expenditure on tuition (1,000s CFA)	0.029	(0.095)	-0.028	(0.119)	0.019	(0.086)	-0.071	(0.121)
Expenditure on supplies (1,000s CFA)	-0.509	(0.314)	-0.529	(0.512)	-0.362	(0.287)	-0.480	(0.396)
Student has tutor	-0.022	(0.028)	-0.005	(0.030)	0.006	(0.030)	-0.004	(0.034)
Home has electricity	0.043	(0.045)	0.041	(0.053)	0.052	(0.045)	0.048	(0.055)
Home has modern toilet	-0.007	(0.042)	-0.015	(0.050)	0.016	(0.045)	0.004	(0.053)
Land owned (hectares)	0.293	(0.504)	0.730	(0.561)	-0.214	(0.322)	-0.155	(0.395)
Interview conducted in French	-0.041*	(0.023)	-0.036	(0.029)	-0.014	(0.025)	0.001	(0.029)

Notes: 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. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$