SERVICE DELIVERY INDICATORS

Data for Results and Accountability

Education service delivery in MOZAMBIQUE

TECHNICAL REPORT

WORLD BANK GROUP

AFRICAN ECONOMIC RESEARCH CONSORTIUM
Consortium pour la Recherche Economique en Afrique

Mozambique 2014 Service Delivery Indicators

Education Technical Report

9/22/2015 10:26 AM

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This report has been prepared in consultation with the Government of Mozambique by Ezequiel Molina and Gayle Martin (Task Team Leaders). The World Bank Team included Owen Ozier, Lucas Arribas Leyton, Mario Vaisman, Brian Stacy, Eleni Papakosta, Jud Shearer, Osvaldo Matavel, Ali Cossing, Tomasia Mataruca, Angelina Come and Agostinho Mucavel, together with the SDI Team members Raihona Atakhodjayeva, Deon Filmer, Christophe Rockmore, and Waly Wane.

Many others helped in various areas of the design, implementation, analysis and dissemination of the Survey, in particular the Mozambique World Bank Education Task Team Leaders Fadila Caillaud, Ana Menezes, and Sophie Naudeau, as well as Bruno Besbas, Rafael Sauce, and Abel Francisco Otacala from the Country Office Team.

Data collection was undertaken by COWI Mozambique consisting of Rachi Picardo (Project Manager), Yara Berti, Ilundi Cabral, Célia Diniz (Team Leaders), and Elias Zavale (Data Base Manager). Further data collection was performed by Feliz Tsandzana, José Souto, and Mário Mungoi (supervisors), Abel Nassone, Alberto Israel, Amira Zacarias, Armado Uamusse, Camacho João, Carla Alberto, Emerson Luis, Esmeraldo Languane, Félix Naife, Fancisca Manjate, Fumo Chacuro, Hélder Sulude, Helder Mazivila, Isac Fernando, Lena Inguane, Jofina Felix, João Muendane, Midália Uamba, Matilde Dimande, Maria Matsinhe, Mariza Artur, Nazareth Mabutana, Nilza Helena, Ricardo Matusse, Rui João, Sonia Gildo, Tarle Aurélio, Tunica João and Yolanda Macie (enumerators), and Ernesto Rodrigues Nhambe, Petra Mondlane, Eduardo Manhique, Helder Devesse, and Reginaldo Fenias (data entry clerks).

The team would like to use this opportunity to thank the officials of the Ministry of Education, in particular, Director Manuel Rego and Jeannette Vogelaar, who helped us along the way. Also, we would like to thank Marie Coastelo Branco, Rachel Waterhouse, Kobi Bentley and Bridget Crumpton from the UK Department for International Development (DFID), who were involved with the project at different stages. Finally, we would also like to thank the World Bank's Mozambique Country Management Unit and, especially, Mark Lundell (Country Director) and Cristina Isabel Panasco Santos (Program Leader) for their guidance and support.

Finally, the team gratefully acknowledges the financial support from the Department for International Development, U.K., DFID, the William and Flora Hewlett Foundation, and the World Bank.

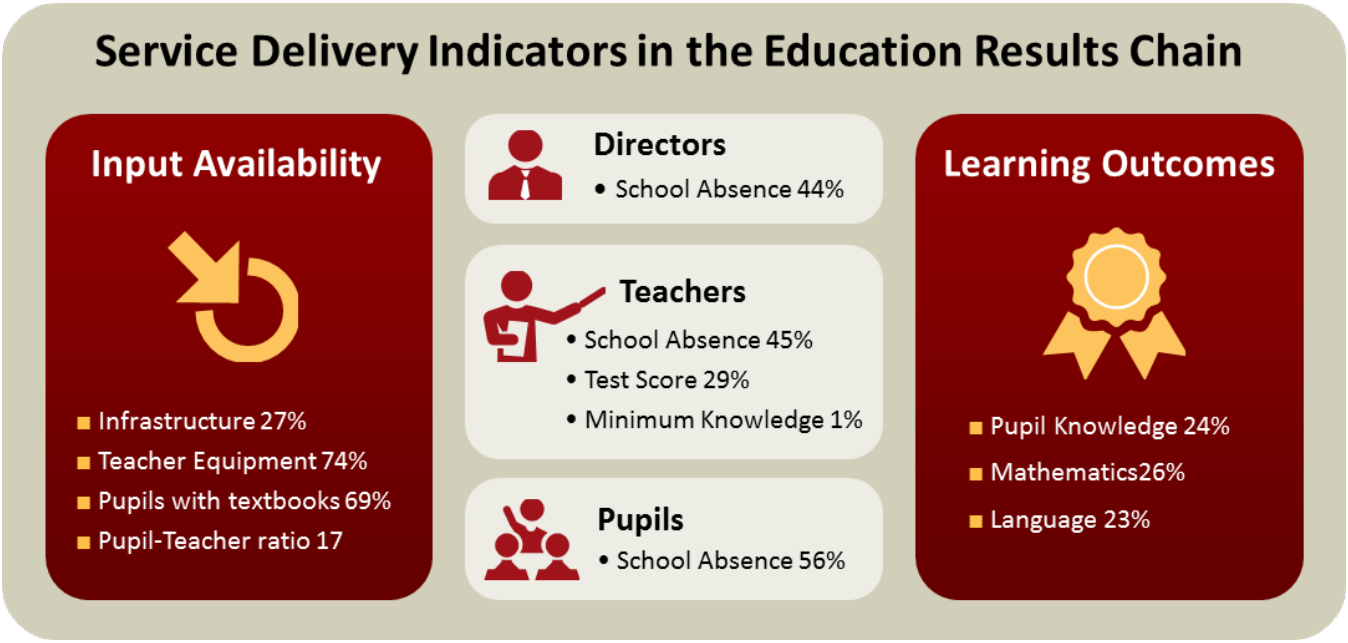
EXECUTIVE SUMMARY

The Service Delivery Indicators provide a set of metrics for benchmarking service delivery performance in education and health. The overall objective of the indicators is to gauge the quality of service delivery in primary education and track performance. The indicators enable governments and service providers to identify gaps and track progress over time and across countries. It is envisaged that the broad availability, high public awareness and a persistent focus on the indicators will mobilize policymakers, citizens, service providers, donors and other stakeholders for action to improve the quality of services and ultimately to improve development outcomes.

This report presents the findings from the implementation of the Service Delivery Indicators in the education sector in Mozambique in 2014. Survey implementation was preceded by extensive consultation with the Government of Mozambique and key stakeholders on survey design, sampling, and adaptation of survey instruments. Pre-testing of the survey instruments, enumerator training, and field-work took place in 2014.

Information was collected from 200 primary public schools, 1,006 teachers and 1,731 pupils. The results provide a snapshot of the quality of service delivery and the physical environment within which services are delivered in public primary schools. The survey provides information on three levels of service delivery: measures of (i) teacher effort; (ii) teacher knowledge and ability; and (iii) the availability of key inputs, such as textbooks, basic teaching equipment, and infrastructure (e.g., sanitation, quality of lighting, etc.). Figure 1 summarizes the results of the survey.

Figure 1. Service Delivery Indicators in the Education Results Chain



Teacher ability: what providers know?

The results indicate that the biggest challenge in Mozambique is the fact that teachers lack the necessary skills to teach effectively. The average score on the Portuguese and mathematics assessment, among Portuguese and mathematics teachers, was 35 percent, with only 0.3 percent of the teachers managing to obtain a score

of at least 80 percent on these assessments. Pupils cannot learn more from their teachers than what the teachers know, and, therefore, teachers' lack of technical competences can constrain learning outcomes.

Table 11 shows selected elements of the teacher assessment by country. In the composition task of the language assessment, for example, teachers in Mozambique only found 10 percent (2 out of 20) errors (such as grammar, punctuation, spelling, syntax, and salutation), compared to an average of 32 percent in other SDI countries (about six out of 20 errors). For example, in mathematics, only 65 percent of teachers in Mozambique could subtract 86-55 compared to the average teacher score of 77 percent of teachers in the other SDI countries.

Teacher effort: what providers do?

On average, 45 percent of teachers were found to be absent from school, which translates into a loss of 1hr 56min of teaching time per day. A further 21 percent of teachers were not in class teaching, leading to another 47 minutes loss of teaching time. When in the classroom, teachers taught 95 percent of the time, translating into 21 minutes more of teaching being lost. Cumulating the sources of lost teaching time, pupils received only 39 percent of the scheduled teaching time. This implies that out of a possible 190 school days, pupils received only 74 effective teaching days, an average of 1hr 41min of teaching per day from a total of 4hrs 17mins they should be receiving.

Availability of key inputs: what providers have to work with?

Compared to the other SDI indicators, Mozambique is doing better in infrastructure availability (such as clean, private available toilets, share of pupils with textbook, and pupil-teacher ratio), but there is room for improvement: The pupil-teacher ratio, as observed during the SDI survey, averaged 21 pupils per teacher, a very low number due to high pupil absence rates. Seventy-seven percent of the schools had the minimum teaching materials and 68 percent of the pupils had a textbook. The minimum infrastructure indicator was low, 29 percent, because of the low percentage of clean toilets in the schools. Only 35 percent of all primary schools surveyed had clean toilets, as opposed to 75 percent who had available toilets and 65 percent with accessible toilets (such as extant and unlocked etc.).

How Mozambique compares with other SDI countries

Mozambique performed poorly relative to the countries where SDI surveys have been implemented (Tanzania, Senegal, Kenya, Uganda, Nigeria, and Togo), particularly in terms of teacher effort and knowledge. Input availability is above the mean of SDI countries. In fact, textbook availability and access to teaching equipment (e.g., blackboard, etc.) is relatively high, and the observed pupil-teacher ratio is the lowest compared to other SDI countries. The performance gaps were especially significant in teacher's school absence where 45 percent of Mozambican teachers were absent from school compared to 30 percent in Uganda, the next worse performer among the other countries. As a result, Mozambique's children get 1hr 41min of teaching time compared to an average of 2hr 53min among other SDI countries. Mozambican teachers also had the lowest overall score (27 points out of 100) in the assessment of their language, mathematics, and pedagogy skills. The next two worst performers, Nigeria and Togo, scored 31 and 34 points, respectively, while, Kenya, the best performing country, averaged 56 points.

In summary, pupils in Mozambique are at a disadvantage in terms of the quality of the education they receive compared to their peers in other SDI countries, which in turn translates into significantly lower learning outcomes evidenced by scores on pupil assessment. In fact, pupils in Mozambique are the lowest performers

among SDI countries, scoring an average 21 out of 100 points compared to an average[1] of almost 45 points in the other SDI countries (Table 1).

[1] Unweighted average.

Table 1. Service Delivery Education Indicators At-a-Glance:

	Mozambique	Urban	Rural	Percent Difference (%)	Southern	Central	Northern
What providers know (ability)							
Minimum knowledge (% of teachers)	0.3	0.0	0.4	-1.0	1.6	0.0	0.0
Test Score (out of 100)	27	25	28	-0.1	32	26	26
What providers do (effort)							
School absence rate (% of teachers)	45	33	48	-0.3***	30	48	53
Classroom absence rate (% of teachers)	56	39	61	-0.4***	41	58	68
Director absence rate (% of directors)	44	30	41	-27	30	49	43
Time spent teaching per day	1hr 41min	2hr 7min	1hr 37min	0.2	2hr 16min	1hr 38min	1hr 20min
What providers have to work with (availability of inputs)							
Observed pupil-teacher ratio	21	40	17	1.3***	33	17	20
Share of pupils with textbooks (% of pupils)	68	57	71	-0.2	66	70	69
Minimum equipment availability (% of schools)	77	92	73	0.3***	92	81	58
(90% with pencils and notebooks)							
Minimum infrastructure availability (% of schools)	29	35	28	0.2	34	36	14
Pupil Learning							
Language and mathematics test score (out of 100)	21	22	21	0.0	37	20	14
Language test score (out of 100)	19	19	19	0.0	38	18	11
Mathematics test score (out of 100)	25	25	25	0.0	29	25	24

Note: Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

Table 2. SDI-At-a-glance (public schools only)

	Mozambique 2014	Average SDI	Kenya 2012	Nigeria[a] 2013	Senegal 2011	Tanzania 2014	Tanzania 2011	Togo 2013	Uganda 2013
Teacher Ability									
Minimum knowledge (% teachers)	0.3	13	35	2.4	Not Comparable	16	Not Comparable	0.9	10
Test score (out of 100)	27	42	56	31	Not Comparable	47	Not Comparable	34	43
Teacher Effort									
School absence rate (% teachers)	45	20	15	17	18	15	23	23	30
Classroom absence rate (% teachers)	56	42	47	23	29	47	53	39	57
Time spent teaching per day	1h 41 min	2h 53min	2h 30min	3h 10min	3h 15min	2h 57min	2h 04min	3h 15min	2h 56min
Scheduled teaching time	*4h 17min*	*5h 31min*	*5h 31min*	*4h 44min*	*4h 36min*	*5h 54min*	*5h 12min*	*5h 28min*	*7h 13min*
Availability of Inputs									
Observed pupil-teacher ratio	21	42	39	22	34	41	74	31	54
Textbook availability (% pupils)	68	37	45	34	Not Comparable	26	Not Comparable	76	6.0
Minimum equipment availability (% classrooms)	77	58	74	48	Not Comparable	62	Not Comparable	24	80
Minimum infrastructure availability (% schools)	29	36	60	13	Not Comparable	36	Not Comparable	14	57
Pupil Learning									
Language and mathematics test score (out of 100)	21	46	69	25	Not Comparable	49	Not Comparable	38	45
Language test score (out of 100)	19	45	73	23	Not Comparable	48	Not Comparable	37	43
Mathematics test score (out of 100)	25	45	57	28	Not Comparable	58	Not Comparable	41	42

Notes: a. Values for Nigeria are the weighted average of the four states surveyed, namely Anambra, Bauchi, Ekiti, and Niger.
b. These numbers reflect the updated SDI methodology. More information can be found on www.SDIndicators.org.
c. Full definitions of indicators in Annex B.

I. INTRODUCTION

The Government of Mozambique (GoM)'s long-term vision to develop basic and life-long education for its citizens is based on the commitment made to the Dakar Declaration *Education for All (2000)* to promote learning of basic skills for all children, young people, and adults, in order to maintain sustainable development and peace, guarantee basic education for all citizens, and reduce illiteracy rates by half by 2015. The GoM is also committed to the Millennium Development Goals (MDGs/OdMs) adopted in 2000 to: 1) eliminate gender disparities in primary and secondary education, by 2015 and 2) completion of primary education for boys and girls by 2015 (Mozambique Ministry of Education, 2012).

The Ministry of Education and Human Development (MoEHD) is extending the *Plano Estratégico da Educação* (PEE) (Education Strategic Plan) until 2019, bringing a stronger focus on further improving education quality, especially during the first few years of primary. The Operational Plan (OP) of the PEE (2015-2018) for primary education identifies four areas of focus for the MoEHD: (i) improving school readiness; (ii) ensuring classroom dynamics that focus on key learning outcomes (basic literacy); (iii) strengthening local governance and responsibility; and (iv) enhancing equitable, effective, and efficient utilization of existing financial and human resources.

The GoM's Poverty Reduction Strategy 2011-2014 (PARP) identifies the following priorities in education: (i) universal access to seven years of primary education of sufficient quality to ensure the learning of basic skills; (ii) expand access for youth and adults to literacy and life skills programs; (iii) expand opportunities for youth and adults to develop meaningful and useful life skills; and (iv) develop and implement mechanisms to give the most vulnerable groups access to the education system. Overall, both access to and quality of education continue to be high priorities in Mozambique. In addition, by focusing on improving learning outcomes, especially for the most vulnerable, the project directly contributes to the World Bank Group's twin goals of ending extreme poverty and boosting shared prosperity.

According to the *Balanco*[2], Mozambique has made significant progress in improving equitable access to primary education, including: 1) increasing the number of classrooms, thus reducing supply side constraints, 2) improving the allocation and availability of school materials, 3) improving equitable access to education, 4) an increase in the number of primary teachers, and 5) a decrease in teachers with no training. Despite considerable progress, participation and learning outcomes in primary school show signs of inequality related to the area of residence and income. Primary education still faces many challenges, particularly in (1) low retention, (2) sub-optimal learning and environment, and (3) poor management at the school level. These elements are all contributing to weak learning outcomes at the lower primary education level.

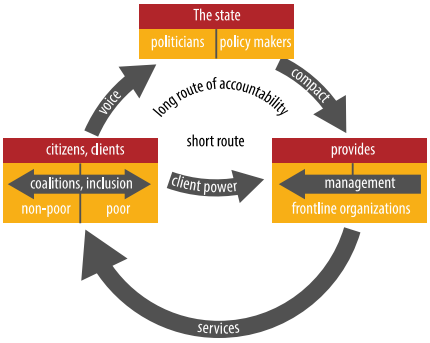
This report presents the findings from the implementation of the first Service Delivery Indicator (SDI) survey in Mozambique. A unique feature of the SDI survey is that it looks at the production of education services at the frontline. The production of education services requires three dimensions of service delivery: (i) providers who exert the necessary effort in applying their knowledge and skills; (ii) providers who are skilled; and (iii) the availability of key inputs, such as equipment and infrastructure. Successful service delivery requires that all these elements are present in the same school and at the same time. While many data sources provide information on the average availability of these elements across the education sector, the SDI survey allows for the assessment of how these elements come together to produce quality education services in the same place at the same time.

[2] The *Balanco* is a document that takes stock of the implementation of the Sector Strategy. It is an internal evaluation done by the Ministry to review the implementation of Sector Strategy.

Box 1. Analytical Underpinnings

Service delivery outcomes are determined by the relationships of accountability between policymakers, service providers, and citizens (Figure 2, World Bank 2004). Human development outcomes are the result of the interaction between various actors in the multi-step service delivery system, and depend on the characteristics and behavior of individuals and households. While delivery of quality education is contingent foremost on what happens in classrooms, a combination of several basic elements have to be present in order for quality services to be accessible and produced by teachers at the frontline, which depend on the overall service delivery system and supply chain. Adequate financing, infrastructure, human resources, material, and equipment need to be made available, while the institutions and governance structure provide incentives for the service providers to perform.

Figure 2. Relationships of accountability between citizens, service providers, and policymakers



Service Delivery Production Function

Consider a service delivery production function, *f*, which maps physical inputs, *x*, the effort put in by the service provider, *e*, as well as his/her type (or knowledge), θ, to deliver quality services into individual level outcomes, *y*. The effort variable, *e,* could be thought of as multidimensional and, thus, include effort (broadly defined) of other actors in the service delivery system. We can think of this type as the characteristic (knowledge) of the individuals who are selected for a specific task. Of course, as noted above, outcomes of this production process are not just affected by the service delivery unit, but also by the actions and behaviors of households, which we denote by ε. We can therefore write:

$$y = f(x,e,\theta) + \varepsilon$$

To assess the quality of services provided, one should ideally measure $f(x,e,\theta)$. Of course, it is notoriously difficult to measure all the arguments that enter the production, and would involve a huge data collection effort. A more feasible approach is, therefore, to focus instead on proxies of the arguments which, to a first-order approximation, have the largest effects.

Indicator Categories and the Selection Criteria

There are a host of data sets available in education. To a large extent, these data sets measure inputs and outcomes/outputs in the service delivery process, mostly from a household perspective. While providing a wealth of information, existing data sources (like Living Standards Measurement Survey (LSMS), Welfare Monitoring Surveys (WMS), and Core Welfare Indicators Questionnaire Survey (CWIQ)) cover only a sub-sample of countries and are, in many cases, outdated.

Box 1. Analytical Underpinnings (cont'd)

The proposed choice of indicators takes its starting point from the recent literature on the economics of education and service delivery, more generally. Overall, this literature stresses the importance of provider behavior and competence in the delivery of health and education services (as opposed to water and sanitation services and housing that rely on very different service delivery models). Conditional on service providers exerting effort, there is also some evidence that the provision of physical resources and infrastructure has important effects on the quality of service delivery.

The somewhat weak relationship between resources and outcomes documented in the literature has been associated with deficiencies in the incentive structure of school and education systems. Indeed, most service delivery systems in developing countries present frontline providers with a set of incentives that negate the impact of pure resource-based policies. Therefore, while resources alone appear to have a limited impact on the quality of education and health in developing countries, it is possible inputs are complementary to changes in incentives, so coupling improvements in both may have large and significant impacts (Hanushek, 2006). As noted by Duflo, Dupas, and Kremer (2011), the fact that budgets have not kept pace with enrollment, leading to large pupil-teacher ratios, overstretched physical infrastructure, and insufficient number of textbooks, etc., is problematic. However, simply increasing the level of resources might not address the quality deficit in education and health without also taking providers' incentives into account.

SDI proposes three sets of indicators: (i) provider effort; (ii) knowledge of service providers and (iii) availability of key infrastructure and inputs at the frontline service provider level. Providing countries with detailed and comparable data on these important dimensions of service delivery is one of the main innovations of the Service Delivery Indicators.

Additional considerations in the selection of indicators are (i) quantitative (to avoid problems of perception biases that limit both cross-country and longitudinal comparisons), (ii) ordinal in nature (to allow within and cross-country comparisons); (iii) robust (in the sense that the methodology used to construct the indicators can be verified and replicated); (iv) actionable; and (v) cost effective to collect.

Table 3. Education indicators

Teacher Effort
School absence rate
Classroom absence rate
Time spent teaching per day
Teacher Knowledge and Ability
Minimum knowledge in Portuguese,
Minimum knowledge in mathematics,
Minimum knowledge in pedagogy
Availability of Inputs
Minimum infrastructure availability
Minimum equipment availability
Share of pupils with textbooks
Observed pupil- teacher ratio

Box 2. The Service Delivery Indicators (SDI) Program

A significant share of public spending on education is transformed to produce good schooling outcomes at schools. Understanding what takes place at these frontline service provision centers is the starting point in establishing where the relationship between public expenditure and outcomes is weak within the service delivery chain. Knowing whether spending is translating into inputs that teachers have to work with (e.g., textbooks in schools), or how much work effort is exerted by teachers (e.g., how likely are they to come to work), and their competency would reveal the weak links in the service delivery chain. Reliable and complete information on these measures is lacking, in general.

To date, there is no robust, standardized set of indicators to measure the quality of services as experienced by the citizen in Africa. Existing indicators tend to be fragmented and focus either on final outcomes or inputs, rather than on the underlying systems that help generate the outcomes or make use of the inputs. In fact, no set of indicators is available for measuring constraints associated with service delivery and the behavior of frontline providers, both of which have a direct impact on the quality of services that citizens are able to access. Without consistent and accurate information on the quality of services, it is difficult for citizens or politicians (the principal) to assess how service providers (the agent) are performing and to take corrective action.

The SDI provides a set of metrics to benchmark the performance of schools and health clinics in Africa. The Indicators can be used to track progress within and across countries over time, and aim to enhance active monitoring of service delivery to increase public accountability and good governance. Ultimately, the goal of this effort is to help policymakers, citizens, service providers, donors, and other stakeholders enhance the quality of services and improve development outcomes.

The perspective adopted by the Indicators is that of citizens accessing a service. The Indicators can thus be viewed as a service delivery report card on education and health care. However, instead of using citizens' perceptions to assess performance, the Indicators assemble objective and quantitative information from a survey of frontline service delivery units, using modules from the Public Expenditure Tracking Survey (PETS), Quantitative Service Delivery Survey (QSDS), and Staff Absence Survey (SAS).

The literature points to the importance of the functioning of schools and more generally, the quality of service delivery. The service delivery literature is, however, clear that, conditional on providers being appropriately skilled and exerting the necessary effort, increased resource flows for health can indeed have beneficial education outcomes.

The SDI initiative is a partnership of the World Bank, the African Economic Research Consortium (AERC), and the African Development Bank to develop and institutionalize the collection of a set of indicators that would gauge the quality of service delivery within and across countries and over time. The ultimate goal is to sharply increase accountability for service delivery across Africa, by offering important advocacy tools for citizens, governments, and donors alike; to work toward the end goal of achieving rapid improvements in the responsiveness and effectiveness of service delivery.

More information on the SDI survey instruments and data, and more generally on the SDI initiative can be found at: www.SDIndicators.org and www.worldbank.org/SDI, or by contacting SDI@worldbank.org.

II. METHODOLOGY AND IMPLEMENTATION

The Mozambique Service Delivery Indicators draw information from a stratified random sample of 200 public schools in urban and rural areas. The full definition of the indicators can be found in Annex B. The details of the sampling strategy are in Annex A, which includes the sampling weighting procedure.

Table 4 provides the details of the sample for the Service Delivery Indicators. In total, 200 primary school were surveyed, including 673 primary school teachers for skills assessment, absence rates among 1,006 teachers and time on task observations from 200 sampled teachers. In addition, learning outcomes were measured for 1,731 grade four pupils. The results presented in this report reflect weighted means using sampling weight and sample design (Annex A).

Table 4. Education SDI sample in Mozambique[3]

Province	Large schools		Small schools		
	Rural	Urban	Rural	Urban	Total
Cabo Delgado	4	1	4	1	10
Gaza	8	1	9	1	19
Inhambane	9	1	9	1	20
Manica	3	1	3	1	8
Maputo City	0	2	0	2	4
Maputo Province	5	1	5	1	12
Nampula	19	2	17	2	40
Niassa	4	1	5	1	11
Sofala	4	1	4	1	10
Tete	4	1	13	1	28
Zambezia	15	2	18	3	38
Total	**84**	**14**	**87**	**15**	**200**

Survey Instruments and Survey Implementation

The survey used a sector-specific questionnaire (Table 5) with all modules administered at the school level. The questionnaires built on previous similar questionnaires based on international good practice for PETS, QSDS, SAS, and observational surveys.

[3] The number of enrolled pupils in fourth grade classes in each school was used to determine if the school was large (equal or above the median) or small (below the median). Depending on the province and the urban/rural location, the median number of pupils enrolled in the fourth grade classes ranged from 25 pupils in rural Maputo province to 322 in urban Maputo province.

Table 5. Education SDI survey instrument

Module	Description
Module 1: School Information	Administered to the head of the school to collect information about school type, facilities, school governance, pupil numbers, and school hours. Includes direct observations of school infrastructure by enumerators.
Module 2a: Teacher Absence and Information	Administered to head teacher and individual teachers to obtain a list of all school teachers, to measure teacher absence, and to collect information about teacher characteristics.
Module 2b: Teacher Absence and Information	Unannounced visit to the school to assess absence rate.
Module 3: School Finances	Administered to the head teacher to collect information about school finances.
Module 4: Classroom Observation	An observation module to assess teaching activities and classroom conditions.
Module 5: Pupil Assessment	A test of pupils to have a measure of pupil learning outcomes in mathematics and language in grade four.
Module 6: Teacher Assessment	A test of teachers covering mathematics and language subject knowledge and teaching skills.

III. RESULTS

A. Teacher effort

The indicators relating to teacher effort (*School absence rate*, *Classroom absence rate,* and *Time spent teaching per day*) are presented in Table 6.

Table 6. Teacher Effort

	All	Urban	Rural	Percent Difference (%)	Southern	Central	Northern
School absence rate (% teachers)	44.8	33.3	47.9	-0.3***	29.6	48.2	53.0
Classroom absence rate (% teachers)	56.2	39.0	60.8	-0.4***	40.9	58.1	67.8
Time spent teaching per day	1h 41 min	2h 7min	1h 37min	0.2	2h 16min	1h 38min	1h 20min
Scheduled teaching time per day[4]	*4h 17min*	*3h 54min*	*4h 22 min*	*-0.1****	*4h 22 min*	*4h 18 min*	*4h 12 min*

Note: Weighted means using sampling weight and the sample design. Results for absence rate based on observations from 1,006 sampled teachers from 200 schools. Results for time on task based on observations from 200 sampled teachers. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

School absence rate

Methodological Note

School absence rate is measured as the share of teachers who are absent from school at the time of an unannounced visit. It is measured in the following way: During the first announced visit, a maximum of ten teachers are randomly selected from the list of all teachers (excludes volunteer and part time teachers) who are on the school roster. The whereabouts of these ten teachers are then verified in the second, unannounced, visit. Teachers found anywhere on the school premises are marked as present.

School absence rate at the national level was high (Table 6), 44.8 percent. Rural schools presented significantly higher absence rates, 47.9 percent, compared to urban schools, 33.3 percent. When looking at absence rates by region, the results show that the South had lower levels of school absence, 29.6 percent, versus the Central and Northern regions, which had much higher absence rates, 48.2 and 53 percent, respectively.

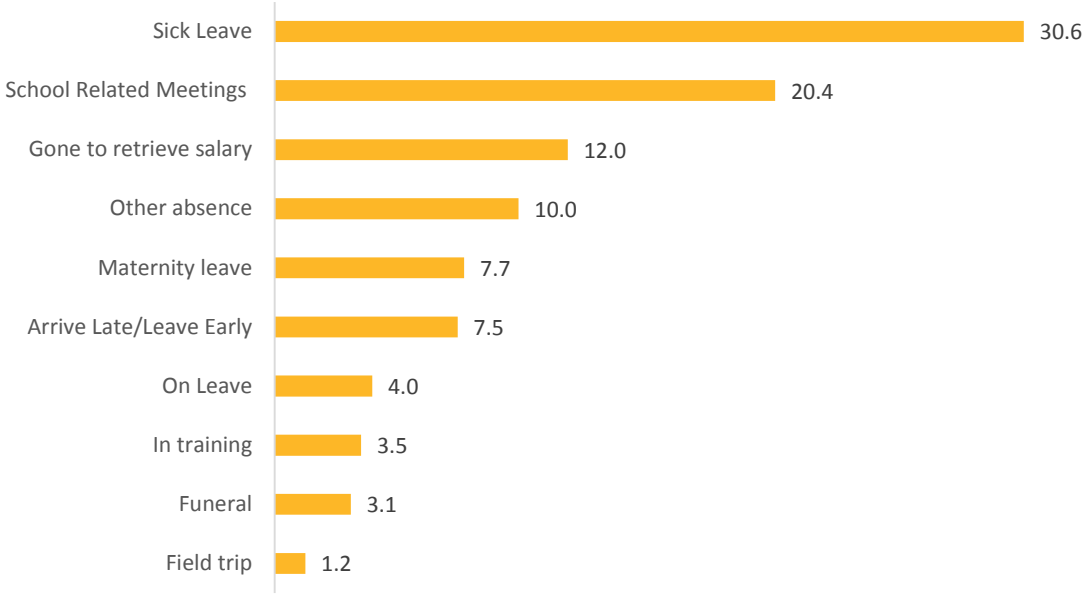
When measuring teacher absence, out of all the teachers, almost 50 percent did not announce they were going to be absent beforehand, and they also did not communicate with the school on that day to explain the reasons for their absence. This means that in any given day 50 percent of all the absences were not approved. When considering all teachers, this means that during any school day a total of 23 percent (0.50*0.45= 0.225) of all teachers in the primary school system were missing from the school without previous notice or approval.

Figure 3 below shows the reasons for absence of the 50 percent of teachers who notified the school of the reasons for their absence as reported by the school director or the teacher in charge. The majority of

[4] *Scheduled teaching time per day* is not an SDI indicator; it is reported in the table for comparison purposes.

reasons for teacher absences (half of them) were due to sick leave (31 percent) or because teachers were attending school related meetings[5] (20 percent).

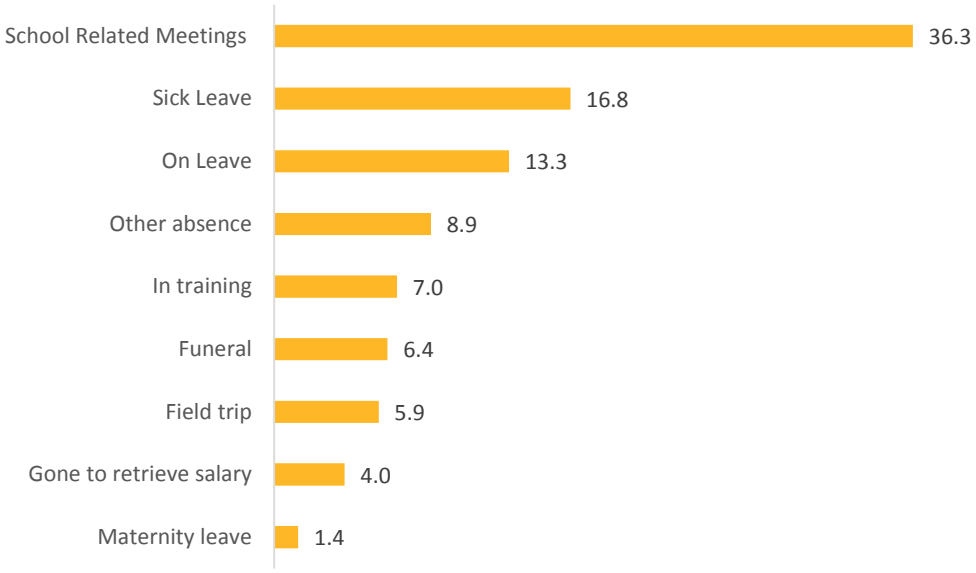
Figure 3. Reasons for teacher absence (percent)



Director absence was calculated the same way as teacher absence. Similar to the teachers' case, almost 50 percent of the directors did not announce their absence beforehand, nor did they communicate on that day the reasons for their absence. Figure 4 shows the reasons for directors' absence for those who notified the school of their absence. Overall, the most frequently reported reason for a director's absence was school-related meetings (36 percent) followed by sick leave (17 percent).

[5] Examples of school-related meetings: meetings of the Zona de Influência Pedagógica (ZIP; Pedagogical Influence Zone) and meetings of the Serviços Distritais de Educação, Juventude e Tecnologia (SDEJT; District Services of Education, Youth and Technology), among others.

Figure 4. Reasons for director absence (percent)



Classroom absence rate

Methodological Note

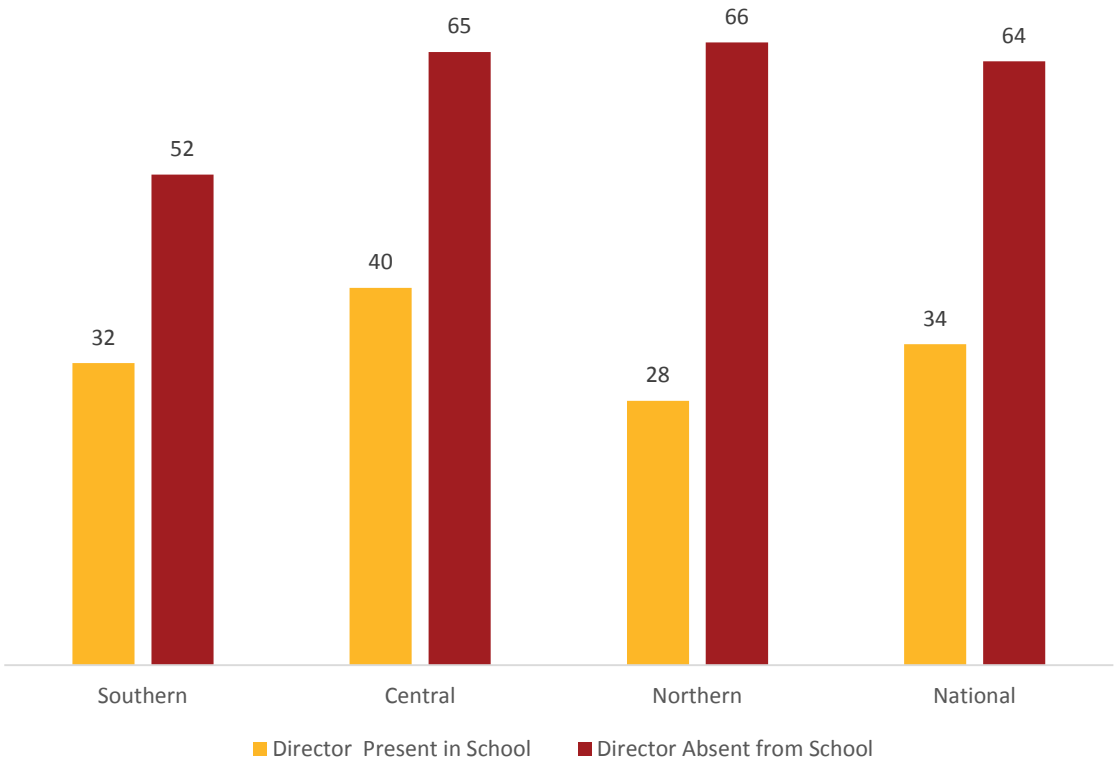
Classroom absence rate is measured as the share of teachers not in the classroom at the time of an unannounced visit. The indicator is constructed in the same way as the school absence rate indicator, with the exception that the numerator now is the number of teachers who are either absent from school, or present at school but absent from the classroom.

Comparing *school* absence rates and *classroom* absence rates, the former appears to be the bigger problem. Almost half of teachers were found to be absent from school, but once teachers were at the school, only 21 percent were not found in class. There were significant differences in classroom absence rates between rural and urban teachers; rural teachers were three times more likely to be absent from class, but at school, compared to teachers in urban areas (25 percent versus eight percent). There was no substantial difference in classroom absence rates between the Southern and the Central regions, 16 and 19 percent respectively. However, classroom absence rates were much higher in the North, which were double compared to the South (32 percent).

Comparison of teacher and director absence rates

Above we documented that there was a high rate of absences that were not approved nor was any notification given to the school. One possible cause of this pattern could be a lack of accountability in the schools. To explore this, we studied whether absence among teachers was related to directors' absence. We found that when a director was present at the school, the average absence rate among teachers was 34 percent. But when a director was absent, the average absence rate for teachers almost doubled to 64 percent (Figure 5). This implies that in schools where the director was absent, teachers were almost twice as likely to be absent, suggesting the importance of director performance in a well-functioning school.

Figure 5. Teacher school absence rate (percent)

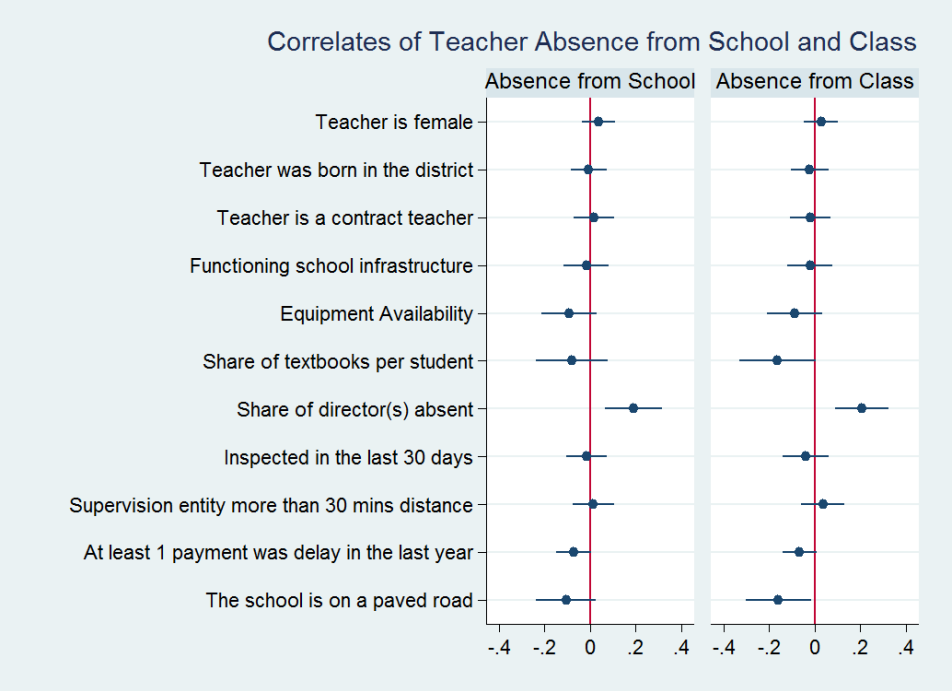


Director presence in school was correlated with teacher presence even after controlling for teacher characteristics (e.g., teacher gender, education level, type of contract, year of experience, whether they was born in the district), school characteristics (school location, infrastructure, equipment, share of pupils with textbooks, access to electricity, whether the school is on a paved road) and management variables (whether the school was inspected in the last 30 days, whether supervision entity was more than 30 minutes of travel time from the school, whether the teacher had received at least one delayed payment in the last year), as well as district fixed effects and dummies for the day of the week in which the survey took place.[6] Figure 6 below shows selected estimates from the regression, as well as the confidence intervals for both dependent variables: absence from the school and absence from class.[7]

Teachers in schools where the director was absent had a 21 percent higher probability of being absent from the school and classroom even after controlling for teacher and school characteristics, as well as external management capability (different from the director).

[6] Standard errors were clustered at the school level.
[7] See Table G 1 in Annex G for the full regression table.

Figure 6. Correlates of teacher absence from school or class



Time spent teaching per day

Methodological Note

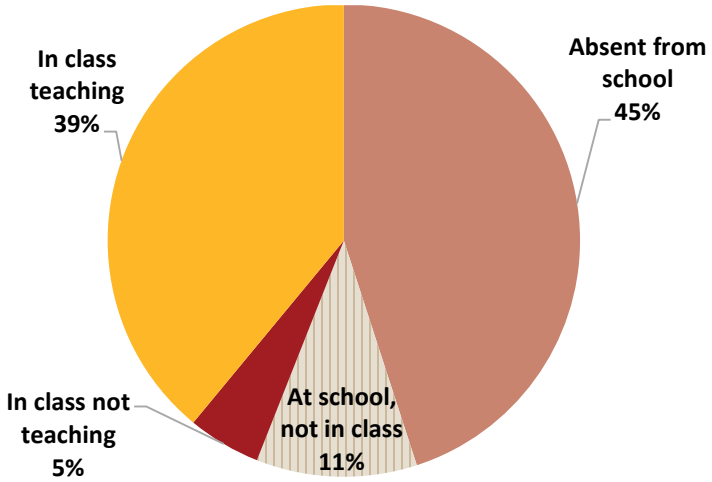
Time spent teaching per day reflects the typical time that teachers spend teaching on an average day. This indicator combines data from the staff roster module (used to measure absence rate), the classroom observation module, and reported teaching hours. The teaching time is adjusted for the time teachers are absent from the classroom, on average, and for the time the teacher teaches while in classroom based on classroom observations. While inside the classroom distinction is made between teaching and non-teaching activities.

Teaching is defined very broadly, including actively interacting with pupils, correcting or grading pupil's work, asking questions, testing, using the blackboard or having pupils working on a specific task, drilling or memorization. Non-teaching activities include working on private matters, maintaining discipline in class or doing nothing and thus leaving pupils not paying attention.

This indicator measures the amount of time a teacher spends teaching during a normal day. The scheduled teaching time for grade four pupils was 4 hours 17 minutes in Mozambique (taking into account break times) (Table 6). It was found that teachers taught only 1 hour 41 minutes. This implies that pupils received on average 74 teaching days out of 190. There was no significant difference between rural and urban schools, while teachers in the South spent approximately 1 hour more teaching compared to teachers in the North. Cumulating this over the school year, it implies that pupils in the North were actually offered 59 days of classes versus 100 days in the South. Figure 7 shows how teachers used the teaching time allocated to them on a regular day.[8] The main leakage came from teachers being absent from school (45 percent).

[8] For the figures, we only took into account the sample of schools where all the components of the teacher time per day were not missing.

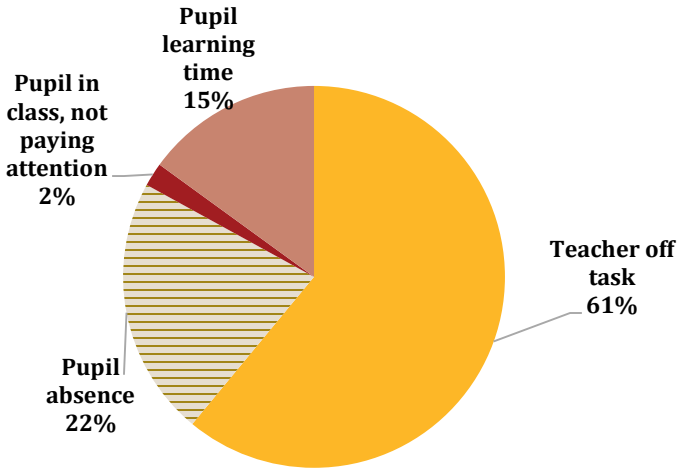
Figure 7. Composition of teaching time spent per day



Note: Teachers Absent from school and those At school, not in class are added up and measured by the indicator Absence from class, 56 percent, (Table 6).

Teachers were in the classroom teaching 39 percent of the time, which means that 61 percent of the time teachers were off task. However, from the pupils' side, we know pupils were present 44 percent of the time and, when in class, they were on task – paying attention to the teacher – 90 percent of the time.[9] Therefore, pupil learning time was 15 percent (0.39*0.44*0.9) (Figure 8). This means that out of a possible 190 school days, pupils received about 29 effective teaching days.

Figure 8. Pupil learning time



[9] For calculations of pupils' absence rate, see the section on input availability. For calculations of pupils off task, see section on Special Focus: Teacher Practices and Classroom Observation.

B. Teacher Competence

Minimum knowledge

Methodological Note

Minimum knowledge is measured as the percentage of teachers who can master the curriculum they taught. It is based on a mathematics and language tests covering the primary curriculum administered at the school level and is calculated as the percentage of teachers who score more than 80% on the language and mathematics portion of the test. The test is given to all mathematics or language teachers that taught 3rd grade last year or 4th grade in the year the survey was conducted.

The objective of the teacher test was to examine the basic reading and writing skills of Portuguese teachers and the arithmetic skills of mathematics teachers that lower primary pupils needed to have to progress further with their education. This was interpreted as the *minimum knowledge* required for the teacher to be effective. Note that the test was validated against the Mozambican primary curriculum, as well as 12 other Sub-Saharan African curricula.[10]

Minimum content knowledge among teachers was shockingly low; only 0.3 percent of teachers scored more than 80 percent on the test (Table 7). There was no significant difference in teacher scores between urban and rural schools. Levels of knowledge were low in both (0.0 versus 0.4 percent). When looking at the data by region, scores in all regions are concerning; the Central (0%) and Northern (0%) regions performed the worst compared to the Southern region, where 1.6 percent of teachers scored above 80 percent on the Portuguese and mathematics test.

Table 7 details the average score on the test and shows the sensitivity of the m*inimum knowledge* indicator to different cut-offs (i.e., requiring a score of 100 percent, 90 percent, and 70 percent). The results were sensitive to the choice of threshold. While there were differences in teacher scores according to the threshold, these differences were not large. When the *minimum knowledge* indicator was calculated as more than 90 percent on the Portuguese and mathematics test, 0.0 percent of teachers had the minimum knowledge, while 1.3 percent of teachers had the minimum knowledge when the threshold was 70 percent. The average score on content knowledge was 38 percent. Table G 2 offers a detailed breakdown of particular questions of each of the three sections of the teacher assessment.

There was a wide variation in the distribution of the teachers' performance on the test (Figure 9), especially for pedagogy. Pedagogical test scores were consistently lower than teachers' language and mathematics test scores, with the largest share of teachers scoring below 20 percent.

[10] See "Teaching Standards and Curriculum Review", prepared as background document for the SDI by David Johnson, Andrew Cunningham, and Rachel Dowling.

Box 3. Assessment of teacher knowledge

Teachers were assessed for their mastery of the primary school level mathematics and Portuguese curriculum, on one hand, and teaching skills, on the other. To test for teacher knowledge in mathematics and Portuguese, teachers were given an indirect test. The test involved asking teachers to mark standardized tasks done by a pupil and suggest a correct answer whenever they indicated the pupil gave the wrong answer. Thus, they were assessed on their ability to identify and suggest a correct answer. The pupil tasks that teachers were asked to mark covered various topics, giving a complete picture of the assessed teachers' mastery of the curriculum.

The test for teaching skills asked teachers to perform tasks they are expected to do to enhance pupil learning, like preparing a lesson plan, evaluating pupils, and tracking progress in pupil performance. For example, teachers were presented with a short story about accidents and asked to prepare a lesson on the reasons road accidents happen and the consequences. Among other things, they were then asked to (i) specify the learning objectives of the lesson; (ii) suggest questions they would ask to determine that pupils understood the lesson and can apply what they have learnt; and (iii) list points for two sides of an argument for a group activity. To test their ability to compare and evaluate pupil performance, teachers were presented with compositions written by two pupils and asked to identify the strengths and weaknesses of each pupil.

Table 7. Teacher assessment

	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Minimum knowledge (% teachers)	**0.3**	**0.0**	**0.4**	**-1**	**1.6**	**0.0**	**0.0**
Language, mathematics and pedagogy average score (out of 100)	26.9	24.6	27.5	-0.1	31.5	25.6	26.3
Language and mathematics average score (out of 100)	34.6	33.8	34.8	0	39.4	33.2	33.8
Minimum knowledge: 100% correct (% teachers)	0.0	0.0	0.0	n/a	0.0	0.0	0.0
Minimum knowledge: 90% correct (% teachers)	0.0	0.0	0.0	n/a	0.0	0.0	0.0
Minimum knowledge: 80% correct (% teachers)	0.3	0.0	0.4	-1	1.6	0.0	0.0
Minimum knowledge: 70% correct (% teachers)	1.3	0.0	1.6	-1.0*	3.4	1.1	0.0

Notes: Weighted means using sampling weight and the sample design. Results for teacher knowledge based on observations from 673 sampled teachers. ^= Minimum knowledge is based on results of the language and mathematics portions of the test. Levels of significance: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

C. Test score

Methodological Note

Test score is measured as the overall score of a language, mathematics and pedagogy tests covering the primary curriculum administered at the school level to all mathematics and language teachers that taught 3rd grade last year or 4th grade in the year the survey was conducted.

Portuguese

Table 8 presents the average score on the language section of the test, as well as a detailed analysis of particular questions. The average score on the Portuguese section was 32 percent, indicating that teachers mastered about one third of the subject matter in the lower primary curriculum. There were no significant differences in language test scores between urban and rural teachers, and only slight differences between regions. Teachers scored much higher in grammar (79 percent), compared to the cloze test[11] (32 percent) and composition (9 percent). When teachers were asked to correct a letter written by a grade four pupil, they only found two out of 20 errors in the letter (such as grammar, punctuation, salutation, spelling, and syntax).

Table 8. Teacher Portuguese assessment

(Out of 100)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Portuguese (complete test)	**32.3**	**34.6**	**31.7**	**0.1**	**36.4**	**33**	**27.9**
Grammar	79	83.2	78.1	0.1	85.5	78.4	75.6
Cloze task	31.5	33	31.1	0.1	35	35.7	20.4
Composition	9.3	11.2	8.9	0.3	12.7	8.7	8.2

Notes: Weighted means using sampling weight and the sample design. Results for teacher knowledge based on observations from 673 sampled teachers. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

Mathematics

Table 9 presents the average teacher scores on mathematics, as well as a detailed analysis of particular questions. The average score on the mathematics section was 30 percent. Rural teachers performed slightly better (31 percent) compared to their urban counterparts (27 percent). Teachers in the Southern region scored better (36 percent) than the Northern (33 percent) and the Central (27 percent) regions.

Looking at the details of the test, 13 percent of teachers could interpret data on a graph, while 25 percent could interpret a Venn diagram. About 12 percent of mathematics teachers could not perform simple tasks such as adding double-digit numbers. Only 17 percent of mathematics teachers were able to compare fractions, and only two thirds of mathematics teachers (60 percent) performed the calculation 86-55 correctly.

[11] A cloze test (also cloze deletion test) is an exercise, test, or assessment consisting of a portion of text with certain words removed (cloze text), where the participant is asked to replace the missing words. Cloze tests require the ability to understand context and vocabulary in order to identify the correct words or type of words that belong in the deleted passages of a text. (http://en.wikipedia.org/wiki/Cloze_test)
Example: Father: Juma, it is (a)_____past seven. Get (b)_____. Today there is a (c)_____football match at school.
Juma: Father, I (d)_____not go to school. I am (e)_____scared to go.

Table 9. Teacher mathematics assessment

(Out of 100)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Mathematics section (complete)	**30.4**	**26.7**	**31.2**	**-0.1**	**36.2**	**27.3**	**32.6**
Lower primary	36.8	32.8	37.7	-0.1	42.7	33.2	40
Upper primary	18.5	15.2	19.2	-0.2	24.1	16.3	18.6
Comparing fractions	16.8	11.4	18	-0.4	16.3	20.1	10.4
Interpreting a Venn diagrams	20.3	22	19.9	0.1	28.3	17.3	20.5
Interpreting data on a graph	12.5	8.1	13.5	-0.4	16.3	11.1	12.8

Notes: Weighted means using sampling weight and the sample design. Results for teacher knowledge based on observations from 673 sampled teachers. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

Pedagogy

The overall score on the pedagogy section of the teachers assessed was 12 percent (Table 10), reflecting difficulties in successfully preparing a lesson plan (17 percent), correctly assessing pupils' writing (10 percent), and using pupils' test scores to make statements about class learning patterns (five percent). The low scores in the pedagogy section, combined with the poor performance on the assessment of curriculum content, imply that teachers had limited content knowledge. These results also indicate that even what the teachers did know, they could not teach adequately.

Table 10. Teacher pedagogy assessment

(Out of 100)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Pedagogy section (complete)	**12.3**	**7.6**	**13.3**	**-0.4*** **	**16.5**	**10.8**	**12.1**
Lesson preparation	16.7	13	17.5	-0.3*	19.8	16.3	15.2
Pupil comparisons	10.4	4.3	11.8	-0.6***	16	7.3	12.5
Pupil evaluations	5.3	1.7	6.1	-0.7***	9	5.1	3.1

Notes: Weighted means using sampling weight and the sample design. Results for teacher knowledge based on observations from 673 sampled teachers. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

SDI Countries

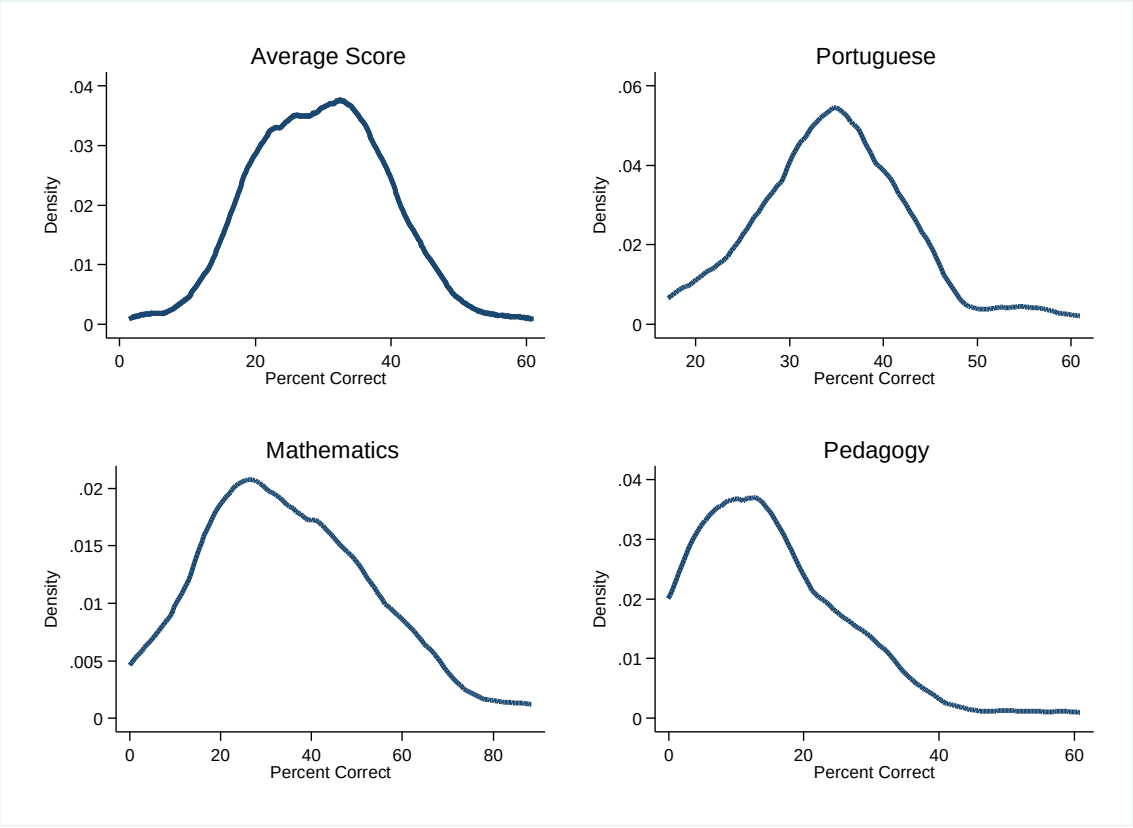
Table 11 shows selected elements of the teacher assessment by country. In the language assessment, for example, teachers in Mozambique only found two out of 20 errors (such as grammar, punctuation spelling, syntax, and salutation), compared to an average of 28 percent in other SDI countries (about six out of 20 errors). In mathematics, only 60 percent of teachers in Mozambique could subtract 86-55 compared to 73 percent, the average teacher score in the other SDI countries. In summary, pupils in Mozambique are at a disadvantage in terms of the quality of the service they receive compared to their peers in other SDI countries, which translates into significantly lower scores on pupil performance. In fact, pupils in Mozambique are the lowest performers among SDI countries, scoring an average 21 percent compared to an average of almost 50 percent in the other SDI countries (Table 16).

Table 11. Teacher competence across SDI countries*

(Out of 100)	Mozambique	Average SDI	Kenya	Nigeria**	Tanzania	Togo	Uganda
Overall Score	**27**	**43**	**57**	**33**	**47**	**36**	**43**
Language (average score)	**32**	**50**	**65**	**46**	**36**	**49**	**51**
Grammar task	70	76	92	60	68	74	88
Composition task	9	30	52	20	17	25	34
Mathematics (average core)	**37**	**57**	**82**	**42**	**65**	**39**	**57**
Adding double digit numbers	82	90	97	84	96	78	95
Subtracting double digits	60	73	87	61	78	64	76
Comparing fractions	17	27	46	13	38	16	20
Subtraction of decimal numbers	36	50	81	41	59	19	50
Pedagogy (average score)	**12**	**26**	**36**	**15**	**35**	**19**	**25**
Preparing a lesson plan	17	34	40	16	58	26	28
Assessing children's abilities	10	25	34	18	17	32	24

Notes: * Data was collected from all schools in the country, public and private; except in the cases of Mozambique and Tanzania where only public schools were surveyed. **Surveyed states in Nigeria were Anambra, Bauchi, Ekiti, and Niger.

Figure 9. Distribution of the teacher test scores



D. Availability

Minimum equipment availability

Methodological Note

Minimum equipment availability is a binary indicator capturing the availability of: (i) functioning blackboard and chalk and (ii) pens, pencils and exercise books in 4th grade classrooms. In one randomly selected 4th grade classroom in the school the enumerator assessed if there was a functioning blackboard by looking at whether text written on the blackboard could be read at the front and back of the classroom, and whether there was chalk available to write on the blackboard. We considered that the classroom met the minimum requirement of pens, pencils and exercise books if both the share of pupils with pen or pencils and the share of pupils with exercise books were above 90%.

The input indicators (*minimum equipment availability*, *minimum infrastructure availability [toilets and light]*, o*bserved pupil-teacher ratio*, and *Share of pupils with textbooks)* were constructed through visual inspections of a grade four classroom and the premises of each primary school.

Table 12. Minimum equipment availability

	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Minimum equipment availability							
Equipment index (% classrooms)	76.8	92.3	73.4	0.3***	91.8	81.0	58.3
Pupils with pencils (% pupils)	96.6	98.5	96.2	0.0**	99.3	97.1	93.8
Pupils with exercise book (% pupils)	97.3	99.7	96.8	0.0***	98.8	98.0	95.0
Classroom with board (% classrooms)	99.0	100.0	98.7	0.0	100	97.9	100.0
Classroom with chalk (% classrooms)	95.6	98.7	94.9	0.0	98.9	100.0	85.4
Contrast to read the board (% classrooms)	91.6	98.3	90.2	0.1**	94.1	94.6	84.4

Note: Weighted means using sampling weight and the sample design. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

The mean outcome for all sub-indicators in the minimum equipment section was high (90 percent or more) in both urban and rural schools; urban schools performed slightly higher in each of these sub-categories (Table 13). Similarly, all three regions performed very well in all sub-indicators, averaging above 90 percent in each sub-indicator with the exception of classroom with chalk (85.4 percent) and contrast to read the board (84.4 percent) in the North.

Minimum infrastructure availability

Methodological Note

Minimum infrastructure availability is a binary indicator capturing the availability of: (i) functioning toilets and (ii) classroom visibility. Functioning toilets is defined as whether toilets were functioning, accessible, clean and private (enclosed and with gender separation) as verified by an enumerator. To verify classroom visibility we randomly selected one 4th grade classroom in which the enumerator placed a printout on the board and checked whether it was possible to read the printout from the back of the classroom.

Share of pupils with textbooks reflects the typical ratio in pupil to textbooks in a 4th grade classroom. It is measured as the number of pupils with the relevant textbooks (language or mathematics conditional on which randomly selected class is observed) in one randomly selected 4th grade class and divided by the number of pupils in that classroom.

Observed pupil-teacher ratio reflects the typical ratio of pupils to teachers in a 4th grade classroom. It is measured as the number of pupils in one randomly selected 4th grade class at the school.

Table 13. Minimum infrastructure availability at the school

	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Functioning school infrastructure (% schools)	**29.1**	**34.5**	**28.0**	**0.2**	**33.6**	**36.0**	**13.8**
Visibility (% classrooms)	97.7	100.0	97.2	0.0**	100.0	99.1	93.6
Toilet available (% schools)	74.7	77.7	74.4	0.0	91.0	77.4	61.7
Toilet clean (% schools)	34.5	49.5	33.0	0.5	35.0	39.2	27.4
Toilet private (% schools)	58.5	66.2	57.7	0.1	76.1	64.2	40.4
Toilet accessible (% schools)	64.5	58.2	65.2	-0.1	80.4	72.1	44.6
Observed pupil-teacher ratio	**21.4**	**40.0**	**17.3**	**1.3*** **	**33.0**	**17.3**	**19.9**
Textbook availability (% pupils)	**68.1**	**56.9**	**70.5**	**-0.1**	**66.3**	**69.6**	**66.8**
Textbook availability (mathematics) (% pupils)	67.8	65.7	67.9	-3.2	62.7	65.9	73.2
Textbook availability (language) (% pupils)	70.8	65.5	71.7	-8.6	76.5	69.6	69.0

Note: Weighted means using sampling weight and the sample design. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

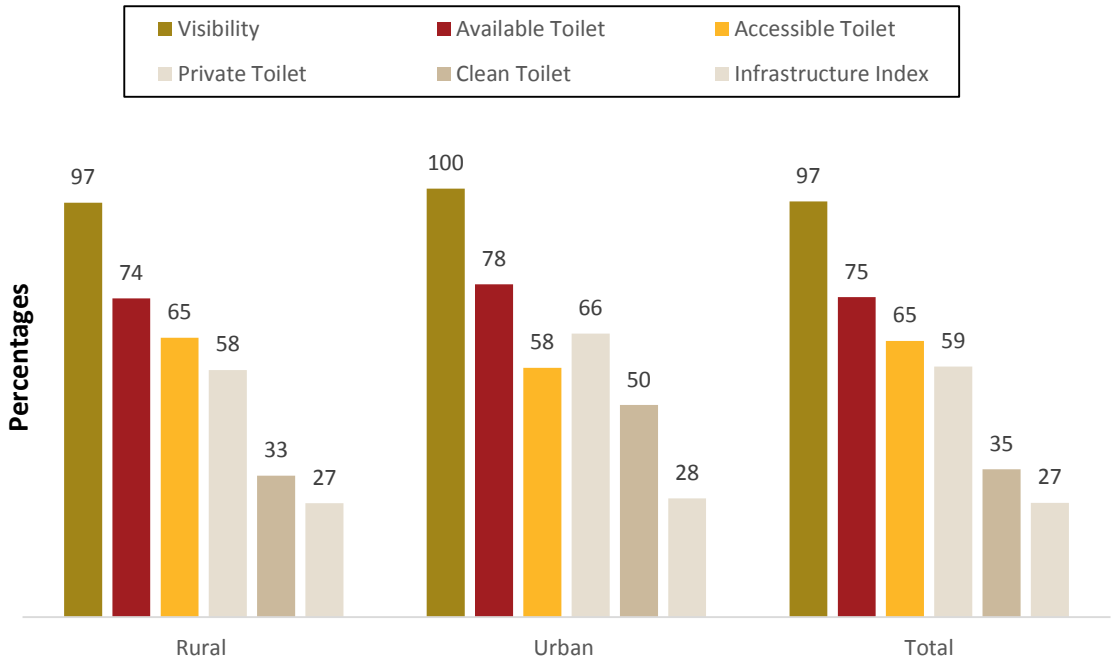
Mozambique did relative well in minimum infrastructure availability (in terms of available toilets and light) compared to other SDI countries (Table 12). However, when looking at functioning toilets (Figure 10), it was observed that even though toilets were available (75 percent), only 35 percent of these toilets were clean, suggesting that infrastructure was available, but not well maintained. The lack of clean toilets, which highlights a human resource problem, was the main constraint in improving infrastructure numbers.

Overall, the Southern region performed better on minimum infrastructure availability, but those differences were not large, as the Central and Northern regions were also doing well. All three regions scored very well in all sub-indicators of the equipment index (e.g., share of pupils with pencils, share of pupils with exercise books, etc.) and relatively well in functioning infrastructure. In regards to available toilets, there was 91 percent toilet availability in the South, 77 percent in the Center, and 62 percent in the North; however, all three regions lagged far behind in the cleanliness of toilets, with only 35 percent clean toilets in the South, 39 percent in the Center, and 27 percent in the North.

Table 12 also reports the share of pupils with textbooks. Overall 68 percent of pupils had a textbook, with more pupils in rural schools having a textbook than in urban schools (71 percent compared to 60 percent). The share of pupils with textbooks in the Central region was higher (70 percent) than pupils in the Southern region (66 percent) and the Northern region (67 percent).[12]

[12] The difference is not statistically significant.

Figure 10. Minimum infrastructure availability



The observed pupil-teacher ratio was low and the lowest among SDI countries; it was at 21.4. Urban schools tended to have larger class sizes (40) compared to rural schools (17.3).[13] There was some variation across schools in Mozambique, with teachers having to teach 33 pupils in the South, while teachers in the Central and Northern regions had to teach 17 and 20 pupils, respectively.

The low numbers of pupils per teacher reflect the problem of pupil absence rates in the schools. While officially there were over 40 pupils enrolled per class, only an average of 21 pupils effectively attended. Pupil absence was highest in the Central (62 percent) and the Northern regions (65 percent) and relatively lower in the Southern region (24 percent).

[13] The difference is statistically significant.

IV. ASSESSMENT OF PUPIL LEARNING

The SDI survey tested grade four pupils. The objective of the pupil assessment was to assess basic reading, writing, and arithmetic skills for fourth grade pupils. The test was designed as a one-on-one test with enumerators reading out instructions to pupils in their mother tongue.

The survey found severe gaps in pupil knowledge. The overall results for all three sections of the pupil assessment (Portuguese, mathematics, and non-verbal reasoning) are reported in Table 14. The table shows the very weak performance of fourth graders in Mozambique. On average, pupils performed poorly, answering correctly only 21 percent of questions on the test. The overall scores were primarily affected by the low scores in language and mathematics. In both tests, pupils from Mozambique were the worst performers in the region (Table 16).

There were small differences in scores between urban and rural pupils, but none of these were significant except for the non-verbal reasoning results, where urban public school pupils scored almost 10 percentage points higher. However, there were noticeable differences in results when comparing regionally. Pupils in the South did better overall (52 percent) than those in the Central and Northern regions (46 percent and 37 percent, respectively).

While the mean score is an important statistic, it is an estimate that by itself is not easy to interpret. Table 14 also highlights the results from the pupil assessment and depicts a more complete breakdown of the Portuguese and mathematics results.[14] In the Portuguese section, 38 percent of the pupils managed the most basic task of reading a letter. However, pupils' scores dropped dramatically as the difficulty increased. Pupils' ability to read a word dropped to 21 percent with the trend continuing as only seven percent could read a paragraph and five percent could correctly answer reading comprehension questions. Regionally, a similar pattern was apparent. The Northern region performed the worst with almost 99 percent of pupils unable to answer reading comprehension questions. Southern pupils scored significantly higher than the other regions in each part of the literacy section of the test. Their scores were double those of the Central region and at least three times higher than pupils in the North. On the mathematics side, scores were slightly higher and some important gaps were revealed. While pupils scored well below 50 percent on all mathematics problems, they were more comfortable with single-digit addition and subtraction. However, they appeared to struggle with any mathematical operation beyond those. Although, it was interesting to notice that pupils' scores in division (8.6 percent) were twice those of their single-digit multiplication skills (4.0 percent). When considering pupils in each region, those in the South performed consistently better than in the rest of the country. Although, the differences in scores were much smaller than in the literacy section.

[14] For a detailed breakdown of the pupil assessment results by gender, see Table G 3 and Table G 4 in Annex G.

Box 4: Background on the SDI pupil assessment

It is instructive to think of the Service Delivery Indicators as measuring key inputs, with a focus on what teachers do and know, in an education production function. These inputs are actionable and they are collected using objective and observational methods at the school level. The outcome in such an education production function is pupil learning achievement. While learning outcomes capture both school-specific inputs (e.g., the quality and effort exerted by the teachers) and various child-specific factors (e.g., innate ability) and household-specific factors (e.g., the demand for education), and thus provide, at best, reduced form evidence on service provision, it is a still an important measure to identify gaps and to track progress in the sector. Moreover, while the Service Delivery Indicators measure inputs -- and learning outcomes are not part of the Indicators -- in the final instance we should be interested in inputs not in and of themselves, but only in as far as they deliver the outcomes we care about. Therefore, as part of the collection of the Service Delivery Indicators in each country, learning outcomes are measured for grade four pupils.

The objective of the pupil assessment was to measure basic reading, writing, and arithmetic skills. The test was designed by experts in international pedagogy and based on a review of primary curriculum materials from 13 African countries (For details on the design of the test, see Johnson, Cunningham and Dowling (2012) "Draft Final Report, Teaching Standards and Curriculum Review"). The pupil assessment also measured nonverbal reasoning skills on the basis of Raven's matrices, a standard IQ measure that is designed to be valid across different cultures. This measure complements the pupil test scores in language and mathematics and can be used as a rough measure to control for innate pupil ability when comparing outcomes across different schools. Thus, the pupil assessment consisted of three parts: language, mathematics and non-verbal reasoning.

The test, using material up to the grade three level was administered to grade four pupils. The reason for choosing pupils in grade four is threefold. First, there is scant information on achievement in lower grades. SACMEQ, for example, tests pupils in grade six. Uwezo is a recent initiative that aims to provide information on pupils' learning irrespective of whether they are enrolled in school or not and tests all children under the age of 16 on grade two material. While this initiative has provided very interesting results, it is not possible to link pupil achievement to school level data, since the survey is done at the household level. Second, the sample of children in school becomes more and more self-selective as one goes higher up due to high drop-out rates. Finally, there is growing evidence that cognitive ability is most malleable at younger ages. It is therefore especially important to get a snapshot of pupil learning and the quality of teaching provided at younger ages.

The test was designed as a one-on-one test with enumerators reading out instructions to pupils in their mother tongue. This was done to build up a differentiated picture of pupils' cognitive skills; i.e. oral one-to-one testing allows us to evaluate whether a child can solve a mathematics problem even when his/her reading ability is so low that he/she would not be able to attempt the problem independently. The language test consisted of a number of different tasks ranging from testing knowledge of the alphabet, to word recognition, to a more challenging reading comprehension test. Altogether, the test included six tasks. The mathematics test also consisted of a number of different tasks ranging from identifying and sequencing numbers, to addition of one- to three-digit numbers, to one- and two-digit subtraction, to single digit multiplication and divisions. The mathematics test included six tasks and a total of 17 questions. The non-verbal reasoning section consisted of four questions.

Table 14. Pupil performance metrics

(Out of 100)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Overall test score (language and mathematics)	**20.8**	**21.6**	**20.7**	**0.0**	**36.9**	**20.0**	**14.4**
Language score	**18.7**	**19.2**	**18.6**	**0.0**	**38.0**	**17.5**	**11.2**
Identify a letter	37.5	35.7	37.8	-0.1	54.1	41.1	24.1
Identify basic words	20.6	20.3	20.6	0.0	44.2	19.6	10.9
Read sentence	12.7	12.3	12.8	0.0	30.7	11.2	6.4
Read paragraph	6.8	7.6	6.6	0.2	18.1	6.3	2.1
Reading comprehension score	5.1	7.3	4.8	0.5	13.3	5.0	1.3
Mathematics score	**25.1**	**25.1**	**25.1**	**0.0**	**28.6**	**25.2**	**23.5**
Single-digit addition	47.9	41.4	48.8	-0.2*	54.9	49.0	42.8
Double-digit addition	18.4	22.0	17.8	0.2	22.9	18.9	15.3
Single-digit subtraction	28.4	29.1	28.3	0.0	38.2	30.7	20.3
Double-digit subtraction	4.6	3.4	4.8	-0.3	6.3	4.5	4.0
Single-digit multiplication	4.0	1.9	4.3	-0.6**	6.9	3.9	2.8
Double-digit multiplication	0.1	0.1	0.1	0.2	0.1	0.1	0.1
Single-digit division	8.6	5.6	9.0	-0.4*	10.3	10.8	4.3
Non-verbal reasoning score	**44.1**	**52.3**	**42.9**	**0.2*** **	**51.8**	**46.0**	**37.3**

Note: Weighted means using sampling weight and the sample design. Results for pupil performance based on observations from 1,731 sampled pupils from 200 schools. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

A regional comparison is depicted in Figure G 1. When looking at the pupil test by region, we found again that the South was doing relatively better than the rest of the country. However, it is important to note that even the Southern region overall was doing worse than any other SDI country. The Northern and Central regions had significantly lower language scores, which might reflect the fact that Portuguese is not widely used for day-to-day communications.

Overall, the average scores on the pupil test were very low and also revealed important areas of the lower primary curriculum that pupils in grade four had not yet mastered (Table 14). In particular, results showed that 62 percent of fourth graders in Mozambique could not identify a letter, less than a quarter could identify words (i.e. *bananas, respect, father, outside*), only 13 percent could read a sentence, and just seven percent could read a paragraph. In mathematics, the results were also worrisome: only half of pupils could do simple addition (e.g., 7+8) and only 18 percent could do double-digit addition (e.g., 27+28). It was clear that basic skills such as these were intended to be taught by grade three. It does not speak well of the match between curriculum goals and pupil achievement that so few pupils in the sample were able to accomplish these grade three tasks when tested halfway through grade four.

The data also revealed that pupils had serious socio-economic impediments: compared to those who did not have breakfast, the pupils who had breakfast before school scored 10 percent higher on the test. A quarter of pupils did not have any breakfast before class and another 25 percent had breakfast with no proteins.

More broadly, the data reveled that pupils in Mozambique suffered from unequal opportunities that affect their performance. Below is an analysis of the links between inequality in outcomes and inequality in opportunities. Inequality in outcomes regards the differences seen across pupils in pupil achievement, while inequality in opportunity regards the gaps in access to high quality education for pupils. Under equality of opportunity, all pupils should have equal access to quality education where a teacher who has mastered the curriculum teaches and the school is well equipped with necessary materials for instruction.

We examined the inequality of outcomes using the pupil achievement measures, and examined inequality of opportunities using the detailed school, teacher, and pupil data. In particular, we formed school ratings based on measures of teacher effort and competency, classroom equipment, and school infrastructure.[15]

While the link we discuss is not necessarily a causal link, we did see a positive association between the school ratings in terms of opportunity and pupil performance.

Figure 11 shows a scatter plot of pupil test scores on school ratings.[16] Pupils in schools on the lower quintile (the worst 25 percent of schools in terms of the quality of the service provided) had a harder time obtaining high scores in the exam, with an average score well below 24 percent, the average for Mozambique as a whole.[17]

[15] The school ratings are based on four components of service delivery: 1. Teacher Effort - measured by the time spent teaching per day; 2. Teacher Competence - measured by average teacher knowledge scores; 3. Classroom Equipment - index combining whether there is a functional blackboard and the share of pupils with pencil and paper and exercise books; 4. School Infrastructure - index combining whether the school has working toilets and sufficient lighting to read the board. The school ratings are an equally weighted average of the four components.

[16] Test scores were aggregated at the school level.

[17] See Table G 5 for more details.

Figure 11. How much do school ratings explain differences in achievement outcomes?

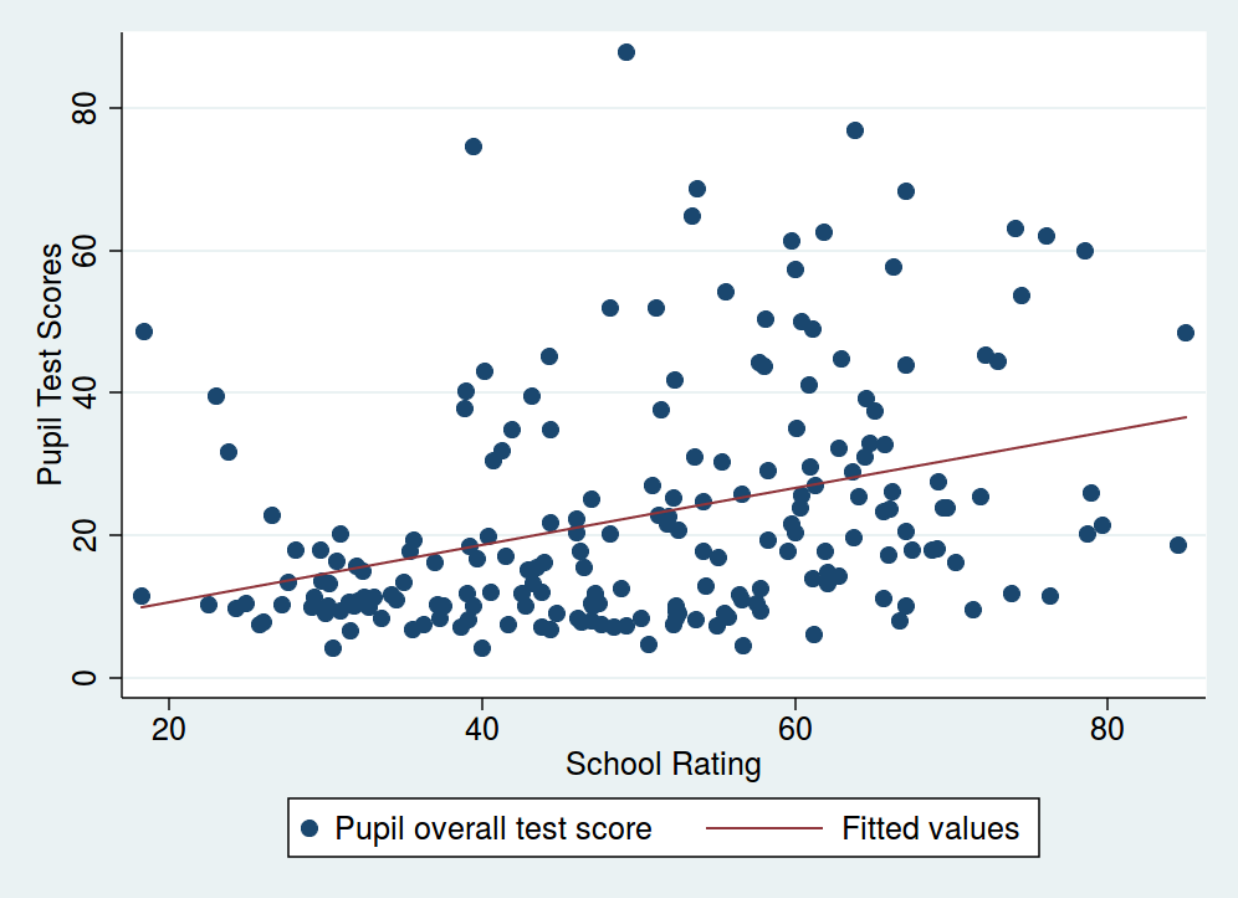


Figure G 2 reveals the distribution in pupil test scores, which clearly shows that in the Portuguese and mathematics sections of the assessment, as well as in their overall performance, the vast majority of pupils scored below 20 percent.

Gender gap evident in the Central and Northern Regions

Figure 12 shows the mathematics and Portuguese test scores for girls and boys by region. Girls from the Southern region were the top performers in language. It is clear that there was no gender gap in the South, but there was in the Central and Northern regions, particularly in Portuguese. In the Central region, boys scored 42 percent higher than girls in Portuguese and 18 percent higher in mathematics. In the North, the situation was similar, with boys scoring 48 percent higher in Portuguese and 24 percent higher in mathematics. These results may, in part, be explained by the fact that the Central and Northern regions have, on average, between 68-80 percent male teachers, while in the Southern region, this percentage goes down to 54 percent.

Figure 12. Pupil performance on language and mathematics by region (percent)

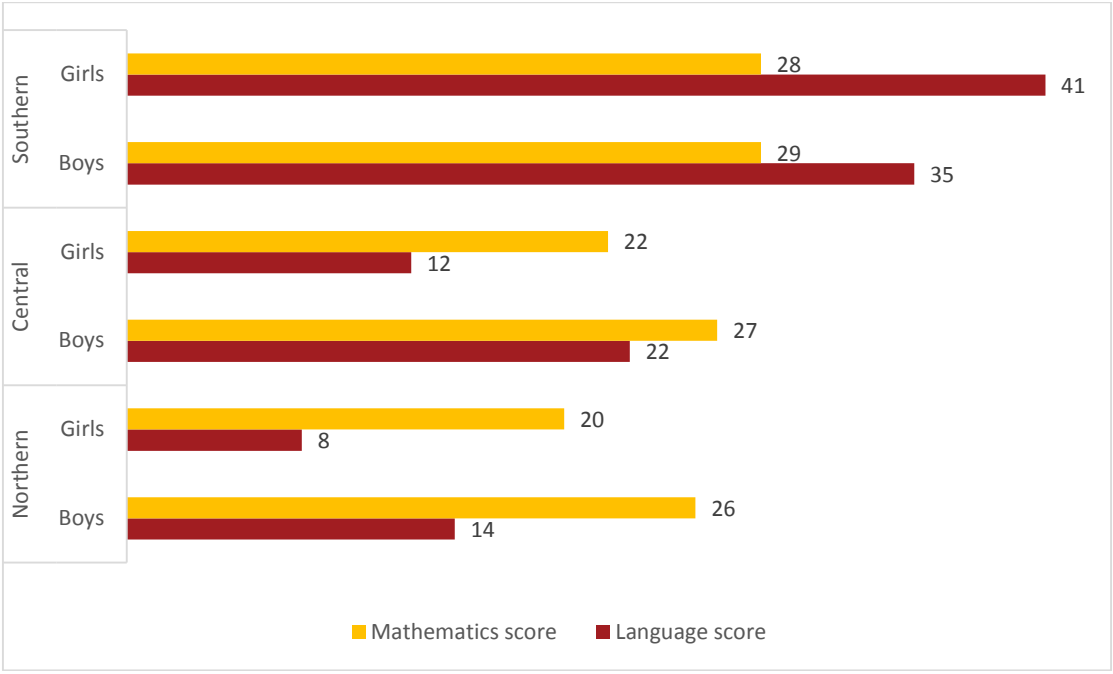


Table 15 shows the gender gap regionally and by geographic location. Nationally, on average, boys outperformed girls in all aspects of the pupil assessment, expect when looking at the Southern region, where girls consistently scored higher in the language assessment. The gap was most noticeable in mathematics, where boys outperformed girls in single-digit addition. In the Central and Northern regions, as well as in urban and rural schools, boys scored between 13.3 and 24.7 percentage points higher in this skill. In the South, boys also scored better, but by a smaller margin (8.9 percentage points). There were similar results in single-digit subtraction, most noticeably with Northern boys scoring 20 percentage points higher than their female classmates. The gender gap was almost nonexistent when considering multiplication of double-digit numbers due to the universally poor results across the country with this skill; the national average for this question type was 0.1 percent for both boys and girls.

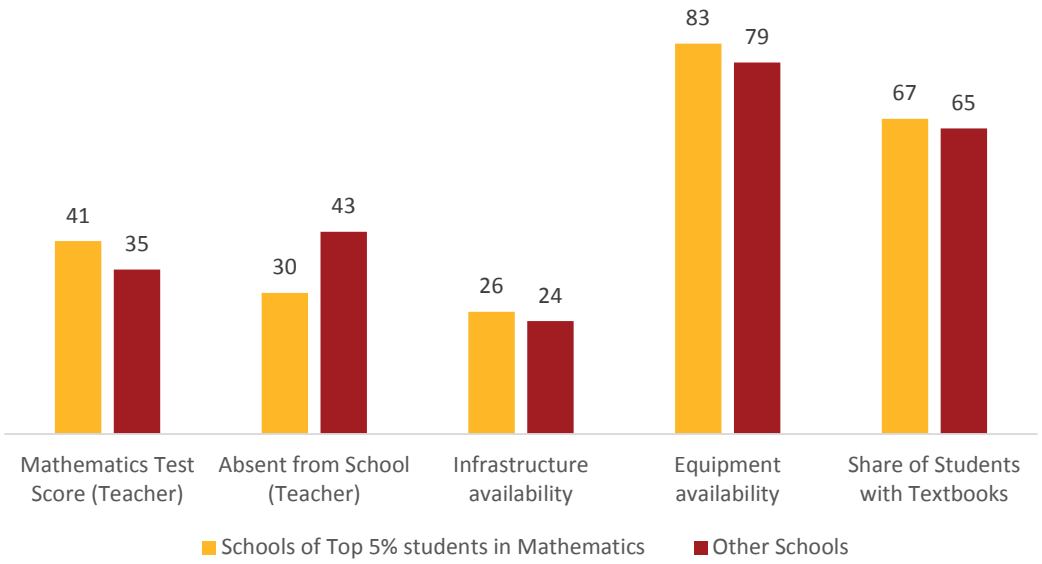
Table 15. Pupil assessment and gender gap

(Out of 100)	All Boys	All Girls	Gender Gap (All)	Gender Gap (Southern)	Gender Gap (Central)	Gender Gap (Northern)	Gender Gap (Urban Public)	Gender Gap (Rural Public)
Overall test score	**22.8**	**18.3**	**4.5**	**-4.7**	**8.7**	**6.1**	**5.1**	**4.6**
Literacy score	**20.8**	**16**	**4.8**	**-6.1**	**9.9**	**6.3**	**5.4**	**4.7**
Identify a letter	40.9	33.3	7.6	2.7	12	5.8	0.5	8.5
Identify basic words	22.8	17.8	5.0	-6.6	9.3	8.9	5.0	5.1
Read sentence	15.2	9.6	5.6	-8.2	11.4	7.5	8.5	5.2
Read paragraph	8.1	5.1	3.0	-4.3	7.2	2.3	5.1	2.7
Comprehension score	5.9	4.1	1.8	-6.0	5.1	2.0	3.0	1.7
Numeracy score	**27.1**	**22.6**	**4.5**	**0.8**	**5.0**	**6.2**	**4.1**	**4.6**
Single-digit addition	54.2	39.9	14.3	8.9	13.2	19.8	13.9	14.0
Double-digit addition	23.3	12.1	11.2	2.3	14.1	12.1	8.1	11.9
Single-digit subtraction	34	21.4	12.6	0.7	14.3	17.5	18.2	11.9
Double-digit subtraction	5.6	3.4	2.2	1.4	1.1	4.9	3.9	1.9
Single-digit multiplication	4.9	2.9	2.0	-2.4	3.7	2.1	0.6	2.1
Double-digit multiplication	0.1	0.1	0.0	0.2	-0.2	0.3	0.3	0.0
Single-digit division	10.3	6.3	4.0	2.1	5.9	2.0	1.9	4.1
Non-verbal reasoning score	**44.3**	**43.8**	**0.5**	**-1.2**	**0.9**	**2.0**	**3.4**	**0.7**

How do schools with the top five percent of pupils differ?

Figure 13 shows the quality of the schools where pupils who scored in the top five percent on the pupil assessment went to school as compared to other schools in the country. The results showed that the schools with the best pupils had teachers that were absent less often (30 percent vs. 43 percent) and had more knowledge (41 percent vs. 35 percent). However, school infrastructure, availability of textbooks, as well as teacher equipment, did not differ significantly between the two groups of schools.

Figure 13. Pupil performance and teacher attendance (percent)



Comparing Mozambique's pupils to other SDI countries

Grade four pupils in Mozambique had the lowest overall performance compared to other SDI countries surveyed (in language, mathematics, and non-verbal reasoning), scoring an average of 21 percent (Table 16). Only 48 percent of pupils could add single digits (e.g., 7+8) and only five percent could subtract double digits (e.g., 57-49). A dismal nine percent and three percent, respectively, were able to successfully perform single- and double-digit division. Kenyan pupils scored seven and 20 times higher, respectively, on these mathematical skills. Pupils in Mozambique scored 19 percent on the language section which is 13 to 57 points lower than in the other countries. In reading comprehension, pupils in other countries at least doubled the scores in Mozambique.

Table 16. Comparison of Pupil Achievement in SDI Countries*

(Out of 100)	Mozambique	Average SDI (unweighted)	Tanzania	Uganda	Togo	Nigeria**	Kenya
Pupil overall score	**21**	**50**	**50**	**49**	**46**	**32**	**72**
Language	**19**	**50**	**48**	**47**	**46**	**31**	**75**
Identify letters	38	79	76	86	78	58	96
Identify basic words	21	57	46	66	65	30	80
Read a sentence	13	47	45	53	26	26	82
Read a paragraph	7	19	24	10	17	12	33
Reading comprehension	5	16	0.8	0.8	18	16	45
Mathematics	**25**	**47**	**58**	**43**	**45**	**32**	**59**
Addition (single digits)	48	77	78	83	77	57	92
Addition (double digits)	18	60	60	56	65	36	84
Subtraction (single digits)	28	70	73	76	65	50	87
Subtraction (double digits)	5	34	38	27	22	22	62
Multiplication (single digits)	4	29	37	24	11	22	51
Multiplication (double digits)	0.1	6	12	2	6	4	8
Division (single digits)	9	38	38	37	36	21	60
Division (double digits)	3	19	21	13	12	12	36
Non-verbal reasoning	**44**	**55**	**54**	**57**	**54**	**50**	**58**
Number of Observations	1731	n/a	4041	3831	1518	6644	2953

Notes: * Data was collected from all schools in the country, public and private; except in the cases of Mozambique and Tanzania where only public schools were surveyed. ** Results for Nigeria are the weighted average of the four states surveyed, namely Anambra, Bauchi, Ekiti, and Niger. Weighted means using sampling weight and the sample design. Results for pupil performance based on observations from 1,731 sampled pupils from 200 schools.

Correlations between indicators and outcomes

The Service Delivery Indicators measure key inputs, with an emphasis on what teachers do and know, in an education production function, and the outcome of this function is pupil learning achievement. With outcome data in education, we can also check whether our input measures are in some ways related to outcomes. Of course, these are mere correlations that do not prove causality. Nevertheless, the focus on SDI only makes sense if they answer the question of how to improve outcomes. Therefore, it is interesting to examine how the Service Delivery Indicators correlate with educational achievement. See Annex D (*Special Focus: Assessing the relationship of what teachers do, what teachers know and pupil performance*) for more details on the relationship between effort, knowledge, and educational outcomes (pupil learning).

Table G 6 depicts unconditional correlations between pupil achievement and the education indicators. Panel A pools school-level data from all schools, while Panel B uses data from the pupil level and controls for the difference between urban and rural schools. Figure G 3 provides a graphical illustration of the correlations between four indicators and pupil test scores.

V. COMPARING MOZAMBIQUE WITH OTHER SDI COUNTRIES

Mozambique performed poorly relative to the countries where SDI surveys have been implemented (Tanzania, Senegal, Kenya, Uganda, Nigeria, and Togo) (Table 17). Textbook availability and access to teaching equipment (e.g., blackboard, etc.) is relatively high, and the observed pupil-teacher ratio is the lowest compared to the other SDI countries. The performance gaps were especially significant in teacher's school absence where 45 percent of Mozambican teachers were absent from school compared to 30 percent in Uganda, the next worse performer. As a result, Mozambique's children get 1hr 41min of teaching time compared to an average of 2hr 53min among the other SDI countries. Mozambican teachers also performed the worst in the teacher assessment. For example, the average combined test score in mathematics, language, and pedagogy was 27 percent compared to the best performing country, Kenya, with a score of 56 percent and the two next lowest performers, Nigeria and Togo, scoring 31 and 34 percent, respectively.

In summary, pupils in Mozambique are at a disadvantage in terms of the quality of the service they receive compared to their peers in other SDI countries, which translates into significantly lower scores on pupil assessment. In fact, pupils in Mozambique are the lowest performers among SDI countries, scoring an average 21 out of 100 points compared to an average of almost 45 points in the other SDI countries (Table 16).

Table 17. SDI-At-a-glance (public schools only)

	Mozambique 2014	Average SDI	Kenya 2012	Nigeria[a] 2013	Senegal 2011	Tanzania 2014	Tanzania 2011	Togo 2013	Uganda 2013
Teacher Ability									
Minimum knowledge (% teachers)	0.3	13	35	2.4	Not Comparable	16	Not Comparable	0.9	10
Test score (out of 100)	27	42	56	31	Not Comparable	47	Not Comparable	34	43
Teacher Effort									
School absence rate (% teachers)	45	20	15	17	18	15	23	23	30
Classroom absence rate (% teachers)	56	42	47	23	29	47	53	39	57
Time spent teaching per day	1h 41 min	2h 53min	2h 30min	3h 10min	3h 15min	2h 57min	2h 04min	3h 15min	2h 56min
Scheduled teaching time	*4h 17min*	*5h 31min*	*5h 31min*	*4h 44min*	*4h 36min*	*5h 54min*	*5h 12min*	*5h 28min*	*7h 13min*
Availability of Inputs									
Observed pupil-teacher ratio	21	42	39	22	34	41	74	31	54
Textbook availability (% pupils)	68	37	45	34	Not Comparable	26	Not Comparable	76	6.0
Minimum equipment availability (% classrooms)	77	58	74	48	Not Comparable	62	Not Comparable	24	80
Minimum infrastructure availability (% schools)	29	36	60	13	Not Comparable	36	Not Comparable	14	57
Pupil Learning									
Language and mathematics test score (out of 100)	21	46	69	25	Not Comparable	49	Not Comparable	38	45
Language test score (out of 100)	19	45	73	23	Not Comparable	48	Not Comparable	37	43
Mathematics test score (out of 100)	25	45	57	28	Not Comparable	58	Not Comparable	41	42

Notes: a. Values for Nigeria are the weighted average of the four states surveyed, namely Anambra, Bauchi, Ekiti, and Niger.
b. These numbers reflect the updated SDI methodology. More information can be found on www.SDIndicators.org.
c. Full definitions of indicators in Annex B.

VI. WHAT DOES THIS MEAN FOR MOZAMBIQUE?

The results provide a snapshot and should not be viewed narrowly as a criticism of teachers, directors or pupils, but as an overview of the state of the education system in the whole of Mozambique. Over time, as the impact of reforms is tracked through repeat surveys in the country, the indicators will allow for tracking of efforts to improve the education system.

There is a clear paradox in the education sector in Mozambique. Increases in both domestic and external investments in the education sector have been significant. As a result, rapid improvements have been seen in terms of availability of inputs (i.e., classrooms, teachers, and learning materials). However, these enhancements have not been translated into better learning outcomes for Mozambican pupils. The 2014 national learning assessment showed that only seven percent of grade three pupils had mastered basic skills. The SDI pupil assessment confirmed this results and shows that pupils struggled with the most basic skills in language (23 percent correct) and mathematics (26 percent correct).

In an attempt to better understand the apparent disconnect between investments and outcomes at the school level, one must look closer at what happens inside the school and inside the classroom. SDI data fills this gap. The SDI data focus on what teacher know, what they do, and what they have to work with. For Mozambique, the data shows that while input availability was at a similar level with other African countries, teacher knowledge nationwide is very low. Only one percent of grade four teachers scored more than 80 percent on the language and mathematics curriculum. This low level of basic knowledge is exemplified by the fact that only 65 percent of mathematics teachers could do double-digit subtraction (e.g., 86-55) and 39 percent could do subtraction with decimals (e.g., 12.15-11.83). In terms of teacher effort, teachers' absence rates were very high; on any given day, 45 percent of teachers were absent from school and 56 percent of teachers were absent from the classroom. Given in other terms, this means that out of 190 schools days, pupils received only 74 effective teaching days. The absence rate of school directors is a matter of concern as well. On average, 44 percent of directors were absent from schools on a daily basis. Unsurprisingly, in schools where the directors were absent, teachers were twice as likely to be absent, as well.

In summary, the transformation of school inputs (i.e., classrooms, teachers, and learning materials) into learning outcomes is impeded by: (i) the low level of ability of teachers in terms of content and pedagogical skills; (ii) the severe deficiencies in service delivery, which is evidenced by the high absence rates of teachers and school directors; and (iii) demand-side constraints which result in high absence rates for pupils.

To remove these obstructions to increase learning levels, the Ministry of Education and Human Development (MINEDH) has identified four areas to be the focus of the MINEDH in primary education, namely: (i) improving school readiness; (ii) ensuring classroom dynamics that focus on key learning outcomes (basic literacy); (iii) strengthening local governance and responsibility; and (iv) enhancing equitable, effective, and efficient utilization of existing financial and human resources.

VII. ANNEXES

ANNEX A. SAMPLING

SUMMARY
The sampling strategy was designed with the dual aims of producing nationally representative estimates and having a minimum power of 80 percent with 0.05 significance level for comparison of key service delivery indicators. The sample strategy also allowed for disaggregation by geographic location (rural/urban) and school size (see Table A 1).

The strata were constructed according to two binary distinctions: urban/rural population; and above- or below-median number of pupils enrolled in the fourth grade by province and urbanization. A total of 200 schools have been sampled and the distribution per province by school type is shown in Table A 1. The sample was reduced from originally 300 schools due to logistical and financial problems, and provide a representative snapshot of the learning environment in public schools in Mozambique.

Table A 1. Sampled schools by geographic location and size

Province	Large schools		Small schools		Total
	Rural	Urban	Rural	Urban	
Cabo Delgado	4	1	4	1	10
Gaza	8	1	9	1	19
Inhambane	9	1	9	1	20
Manica	3	1	3	1	8
Maputo City	0	2	0	2	4
Maputo Province	5	1	5	1	12
Nampula	19	2	17	2	40
Niassa	4	1	5	1	11
Sofala	4	1	4	1	10
Tete	4	1	13	1	28
Zambezia	15	2	18	3	38
Total	**84**	**14**	**87**	**15**	**200**

Sampling Frame

The target population in the education survey was all public primary-level school children. Since parts of the school questionnaire were to be administered to teachers and pupils at the grade four level, all public schools with at least one grade four class formed the sampling frame. We decided to eliminate private schools from the sample frame, as there were only 80 in all of Mozambique, which represented less than one percent of the total sample, both in term of actual schools as well as number of grade four pupils. Similar reasoning led us to also eliminate from consideration community schools, special schools (for pupils with special needs), and schools with only night classes. We used the March 3rd school census in developing the sampling frame for Mozambique, as it was the most complete listing available.

The sample was restricted to schools that had at least one fourth grade class, which eliminated around 300 schools from the sample frame (Table A 2).

Table A 2. Sample Frame - Schools by Urbanization and Size

Province	Large schools		Small schools		Total
	Rural	Urban	Rural	Urban	
Cabo Delgado	420	12	437	13	882
Gaza	332	14	345	15	706
Inhambane	353	27	368	27	775
Manica	340	14	351	15	720
Maputo City	0	51	0	51	102
Maputo Province	193	31	198	31	453
Nampula	716	54	1053	54	1877
Niassa	433	12	466	12	923
Sofala	364	25	382	25	796
Tete	496	16	501	16	1029
Zambezia	1190	108	1461	116	2875
Total	**4837**	**364**	**5562**	**375**	**11138**

Stratification

The schools were disaggregated by sub-national strata (provinces), urban/rural location and size (large/small). In the case of Mozambique, there were three regions (Northern, Central, and Southern), 11 provinces, and 148 districts. Based on the most recent available data from national statistical authority, each of the districts were categorized as rural or urban, as well as each school within them. To classify schools as small or large, we used the size of the grade four class (in terms of the number of pupils enrolled in 2013) in relation to the median by province and urbanization.[18] These distinctions yielded 42 strata from which to sample schools.[19]

Sampling Strategy

The sampling strategy was a simple random sample using the stratification detailed above. However, during fieldwork, the sample was reduced from originally 300 schools to 200 schools due to logistical and financial problems. The sample was originally drawn for those 300 schools and the necessary sample reduction was done after 113 schools had already been surveyed. To decide which additional 87 schools were to be sampled, we used the following decision criteria: For each open stratum, we decided to close the stratum if the resulting standard error was lower than the resulting standard error from randomly sampling using the above mentioned methodology.

$$\text{SE(Close Open Stratums)} < \text{SE(Sample Randomly)}$$

$$\frac{K^2}{N^2}\frac{1}{S_K} + \frac{(N-K)^2}{N^2}\frac{1}{87-(S_k-V_k)} < \frac{1}{87}$$

where K was the population in the stratum, N was the sum of total population in remaining open strata, S_K was the sample for the strata k, and V_k was the number of visited schools in the strata k.

Using this formula, we decided to close four open stratums containing 23 schools. The remaining 64

[18] Three schools were missing information on the size of grade four, and were treated as "large." These were: EPC 26 de Septembro (ex-Boleia) in the Angoche province of Nampula; Escola Primaria de Cunhuma in Cuamba, Niassa; and Escola Primaria de Ualamua in Mauea, Niassa.

[19] Depending on the province and the urban/rural location, the median number of pupils enrolled in the fourth grade classes ranged from 25 pupils in rural Maputo province to 322 in urban Maputo province.

schools were chosen randomly.

Replacement schools were drawn from each location in case the sampling frame included schools that no longer existed, were not functional, or were inaccessible due to security concerns. Note, these back-up facilities were not to be used for logistical ease. Replacement schools were selected in keeping with the probability sampling approach.

Sampling Strategy in Detail[20]

This section explains in detail the methodology used to select which schools to choose from once we reduced the sample size from 300 to 200 schools, as well as how to re-construct the weights.

Part 1: Definitions

Basic variable definitions:

Total population:	N_{total}
Population in one of the strata:	K (or K_i)
Sampled number in one of the strata:	S (or S_i)
Outcome in facility j within stratum i:	X_{ij}
Variance of X_{ij} (ignoring stratum effects):	σ^2

After initial fieldwork, some strata were "closed," meaning that all S_i of K_i had been visited. Others were "open," meaning that some had been visited, and some had not. In this setting, we made three further definitions:

Number visited during initial fieldwork in one of the strata:	V (or V_i)
Total population across all "open" strata:	N (or N_i)
Additional ("top-up") number to visit in one of the strata:	t (or t_i)

Part 2: In-field sampling problems

The scenario we faced was that the budget constraint became tighter during field work, so fewer facilities were able to be visited than originally envisioned. We sampled 300 schools, but could only go to 200. We learned this after visiting 112 and closing 15 out of the 42 strata. Based on this, we needed to visit 88 more schools. The question was: which ones? The decision had to be made very quickly, so we made simplifying assumptions in order to rapidly develop a strategy.

A stark view of the situation held that any sub-sample that had been visited already was subject to potential bias towards "easy-to-visit" facilities that might have different observable and unobservable characteristics from others. The options, in each stratum, were then to re-sample afresh, or to "close" the stratum by completing the visits to the originally sampled facilities.

If we re-sampled all "open" strata afresh, *completely ignoring work that had been done*, we would need to visit 87 facilities, and these facilities would be drawn in a way that would represent the total "open" population. Simplifying away stratum-fixed effects, then, in relation to the variance, σ^2, of the underlying random variables of interest, X_{ij}, the variance of the resulting estimator would be:

[20] This section was written by Owen Ozier, DECRG.

$$Var\left(\frac{1}{88}\sum_{j=1}^{88} X_{ij}\right) = \frac{1}{88}\sigma^2$$

If, instead, a single stratum under consideration were "closed," visiting the remaining sampled facilities, leaving the rest of the open strata to be randomly re-sampled in this way, we would have the following estimator:

$$\frac{K_1}{N}\cdot\frac{1}{S_1}\sum_{j=1}^{S_1} X_{1j} \;+\; \frac{N-K_1}{N}\left(\frac{1}{88-(S_1-V_1)}\sum_{j=1;i\neq 1}^{88-S_1} X_{ij}\right)$$

The variance here was:

$$\left(\frac{K_1^2}{N^2}\frac{1}{S_1} \;+\frac{(N-K_1)^2}{N^2}\frac{1}{88-(S_1-V_1)}\right)\sigma^2$$

The decision was over which variance was greater. Dividing out by the common variance term, in an earlier write-up with slightly different notation, this decision was given by the following equation:

$$\text{SE(Close Open Stratum)} < \text{SE(Sample Randomly)}$$

$$\frac{K^2}{N^2}\frac{1}{S_K} \;+\frac{(N-K)^2}{N^2}\frac{1}{88-(S_k-V_k)} < \frac{1}{88}$$

For large *S* and small *V*, this clearly favored closing the open stratum. In the opposite conditions, it did not. Once the decision was made for one stratum to be closed, the problem could be re-optimized for the remaining strata. We did not check whether this algorithm was globally optimal, but it would lead to a local optimum. Four strata were closed containing 13 schools, leaving 75 randomly sampled from the remaining open strata (23).

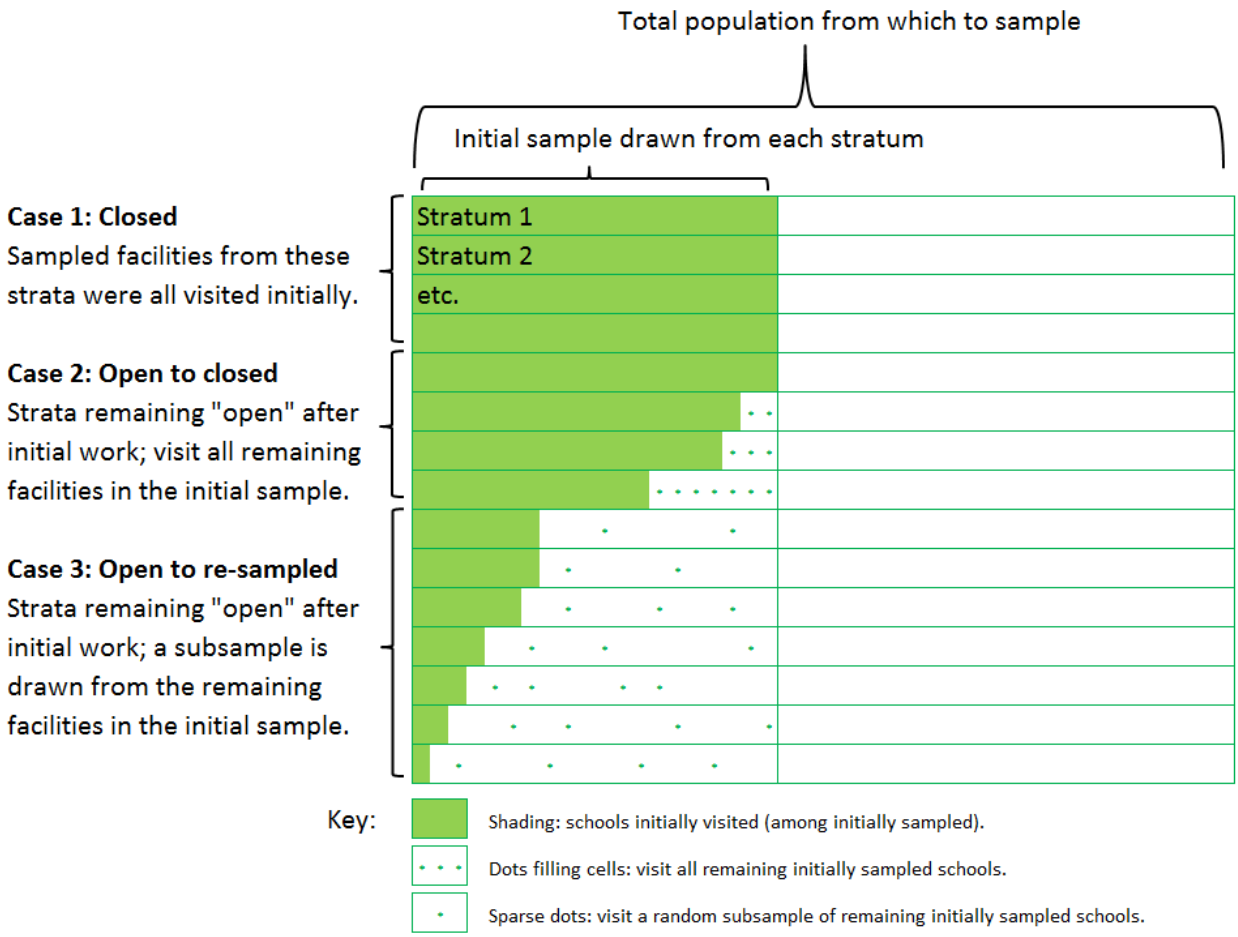
Part 3: How to weight observations in the analysis.

A simplifying assumption we made was that for open strata that we didn't close, we would re-sample afresh, ignoring work done previously. Procedurally, this meant that we didn't need to revisit those re-sampled facilities that *happened* to have already been visited. However, we would have to visit the rest of the re-sampled schools and could discard the observations of the others. The "top-up" number of new visits needed in a given stratum would then fall between 0 and all the remaining facilities:

$$0 < t_i \leq (S_i - V_i).$$

In reality, we intended to use the data already collected, and to re-sample only from the $(S_i - V_i)$ facilities that had been originally sampled but not yet visited in each stratum i. See Figure A 1 below:

Figure A 1.Understanding the Sampling Strategy



The next question we faced was how to weight the resulting facilities in the estimation process.

The goal was to weight each observation by the inverse of the probability that it was included in the ultimate sample. This meant that the original plan called for weighting each observation by the stratum size over the sampled number in that stratum:

$$\text{Original weights: } (K_i \,/\, S_i)$$

In Cases 1 and 2 shown in Figure 14, the original strategy and weights still held. The question was what to do in Case 3.

The simple approach to Case 3, though it was not preferred, was to take the probability to be fixed at the stratum level. Had the initial visits been randomized within each stratum sample, this would be true. The new inverse probability weight would then be:

$$(K_i \,/\, V_i \; + t_i)$$

We suspected that the initially visited facilities were systematically different – they were easy to visit, and may have been different in both econometrically observable and unobservable ways.

They were likely not representative of the strata. However, we proposed this approximation that we believed handled Case 3 more accurately:

The initially visited V_i of the initially sampled S_i facilities in a stratum of total size K_i were representative of a fraction of the stratum: in particular, they were representative of a fraction approximately equal to V_i/S_i of the stratum. As such, for those facilities that were initially visited, the V_i initially visited facilities retain their original weights, (K_i / S_i).

The top-up sample in each stratum, totaling t_i, represented (approximately) the rest of the stratum: a fraction of approximately $(S_i - V_i)/S_i$ of the stratum. Since the stratum was of total size K_i, the right weight – the inverse of the probability that one of the newly re-sampled t_i facilities was ultimately visited – was given by:

$$((S_i - V_i) \cdot K_i) \; / \; (S_i \cdot t_i)$$

In practice, the two weighting schemes described above may or may not yield very different estimated means or standard errors, depending on the actual numbers in the sample and the extent and nature of heteroskedasticity in the observations. One check to ask whether those facilities visited earlier were not representative along observables, was simply to estimate the following equation using OLS:

$$Y_{ij} = \beta \, \text{Early}_{ij} + \Sigma_i \, \gamma_i \, \text{Stratum}_i + \varepsilon_{ij}$$

Here, Early_i was an indicator for whether a facility was visited before (1) or after (0) the fieldwork was halted for re-sampling; the set of indicators Stratum_i were used to estimate fixed effects for strata; and Y_i was any outcome or observable of interest.

ANNEX B. DEFINITION OF INDICATORS

School absence rate	
Share of a maximum of 10 randomly selected teachers absent from school during an unannounced visit	It is measured as the share of teachers who are absent from school at a time of an unannounced visit. It is measured in the following way: During the first announced visit, a maximum of ten teachers are randomly selected from the list of all teachers (excludes volunteer and part time teachers) who are on the school roster. The whereabouts of these ten teachers are then verified in the second, unannounced, visit. Teachers found anywhere on the school premises are marked as present.
Classroom absence rate	
Share of teachers who are present in the classroom during scheduled teaching hours as observed during an unannounced visit	The indicator is measured as the share of teachers not in the classroom at the time of an unannounced visit. The indicator is constructed in the same way as school absence rate indicator, with the exception that the numerator now is the number of teachers who are either absent from school, or present at school but absent from the classroom.
Time spent teaching per day	
Amount of time a teacher spends teaching during a school day	This indicator reflects the typical time that teachers spends teaching on an average day. This indicator combines data from the staff roster module (used to measure absence rate), the classroom observation module, and reported teaching hours. The teaching time is adjusted for the time teachers are absent from the classroom, on average, and for the time the teacher teaches while in classrooms based on classroom observations. While inside the classroom distinction is made between teaching and non-teaching activities. Teaching is defined very broadly, including actively interacting with students, correcting or grading student's work, asking questions, testing, using the blackboard or having students working on a specific task, drilling or memorization. Non-teaching activities includes working on private matters, maintaining discipline in class or doing nothing and thus leaving students not paying attention.
Minimum knowledge	
Share of teachers with minimum knowledge	It is measured as the percentage of teachers who can master the curriculum they taught. It is based on mathematics and language tests covering the primary curriculum administered at the school and is calculated as the percentage of teacher who score more than 80% on the language and mathematics portion of the test. The test is given to all mathematics or language teachers that taught 3rd grade last year or 4th grade in the year the survey was conducted.
Test score	It is measured as the overall score of a mathematics, language and pedagogy tests covering the primary curriculum administered at the school level to all mathematics and language teachers that taught 3rd grade last year or 4th grade in the year the survey was conducted.

Minimum infrastructure availability	
Unweighted average of the proportion of schools with the following available: functioning electricity and sanitation	It is a binary indicator capturing availability of: (i) functioning toilets and (ii) classroom visibility. Functioning toilets is defined as whether toilets were functioning, accessible, clean and private (enclosed and with gender separation) as verified by an enumerator. To verify classroom visibility we randomly select one 4th grade classroom in which the enumerator places a printout on the board and checks whether it was possible to read the printout from the back of the classroom
Minimum equipment availability	
Unweighted average of the proportion of schools with the following available: functioning blackboard with chalk, pens or pencils, and notebooks or paper	It is a binary indicator capturing availability of: (i) functioning blackboard and chalk and (ii) pens, pencils and exercise books9 in 4th grade classrooms. In one randomly selected 4th grade classroom in the school the enumerator assessed if there was a functioning blackboard by looking at whether text written on the blackboard could be read at the front and back of the classroom, and whether there was chalk available to write on the blackboard. We considered that the classroom meet the minimum requirement of pens, pencils and exercise books if both the share of students with pen or pencils and the share of students with exercise books are above 90%.
Share of pupils with textbooks	
Number of mathematics and language books used in a grade four classroom divided by the number of pupils present in the classroom	The indicator reflect the typical ratio in student to textbooks in the 4th grade classroom. It is measured as the number of students with the relevant textbooks (mathematic or language conditional on which randomly selected class is observed) in one randomly selected 4th grade class and divided by the number of students in that classroom.
Observed pupil-teacher ratio	
Average number of grade four pupils per grade four teacher	This indicator reflects the typical ratio in pupils to teachers in the 4th grade classroom. It is measured as the number of students in one randomly selected 4th grade class at the school.

ANNEX C. SPECIAL FOCUS: TEACHER PRACTICES AND CLASSROOM OBSERVATION

Table C 1 describes the sample for the classroom observations. Two hundred classrooms were observed (118 mathematics and 82 Portuguese classes). The sample is balanced across regions, with all regions having between 55 and 61 percent of the sampled mathematics classes and the rest (39 to 45 percent) of Portuguese classes.

Table C 1. Descriptive statistics on sample

Region	Subject					
	Mathematics		Portuguese		Total	
	No.	Row %	No.	Row %	No.	Row %
Southern	20	54.8	17	45.2	37	100.0
Central	58	59.6	39	40.4	97	100.0
Northern	40	60.9	26	39.1	66	100.0
Total	118	59.2	82	40.8	200	100.0

Classroom Observation and Teacher's Time Use

On average, enumerators observed almost 45 minutes of class, which was the allocated time for one lesson in Mozambique. Figure C 1 shows how teachers used the classroom time in one minute intervals. In a typical class, teachers spent the first quarter of the class presenting the lesson (interacting with a large group of pupils), writing on the board, and then waiting for pupils to copy notes in their notepads or exercise books. From this, a picture began to emerge showing what happens during the class. However, additional information was needed to be able to distinguish between various classroom interactions. For example, more analysis was required to differentiate between a teacher that was reprimanding a pupil for a wrong answer when interacting one-on-one and one that was correcting a pupil and encouraging her to try again. In the next section, we will explore exactly what practices were used (i.e., introducing and summarizing the lesson, making the pupils use the textbooks, giving pupils' tasks to practice what they learned, etc.).

Figure C 1. Teacher Use of Classroom Time by Minute

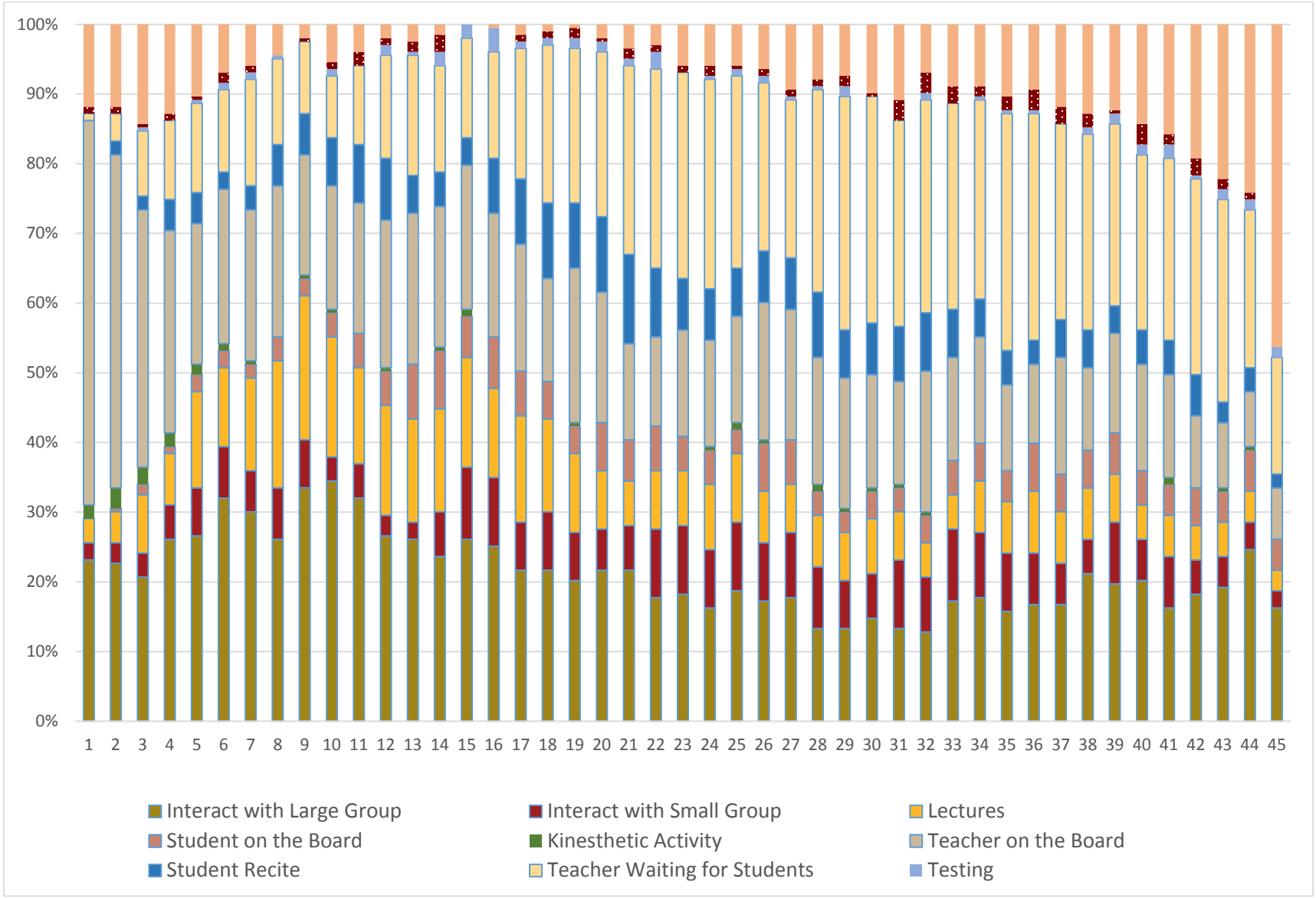


Figure C 2 shows another important aspect of how time was used in the classroom. In this case, rather than looking at the teacher, we look at the pupils. Teaching is hard, especially with fourth graders. If teachers do not motivate pupils to pay attention, the activities or practices the teachers are doing will not affect pupils' understanding of the material, nor will it help increase their knowledge. By measuring every five minutes, the number of pupils that were off task and not paying attention to the lesson, we created a picture of the ratio of pupils who were not paying attention in a given class. The number was lower for the Southern region (eight percent during the Portuguese lesson and five percent during mathematics), a little higher in the Central region (seven percent and nine percent, respectively), and substantially higher in the Northern region (16 percent and 14 percent, respectively). The overall number for Mozambique in terms of pupils who were not paying attention in class was 10 percent.

Figure C 2. Percentage of pupils off-task by subject and region (percent)

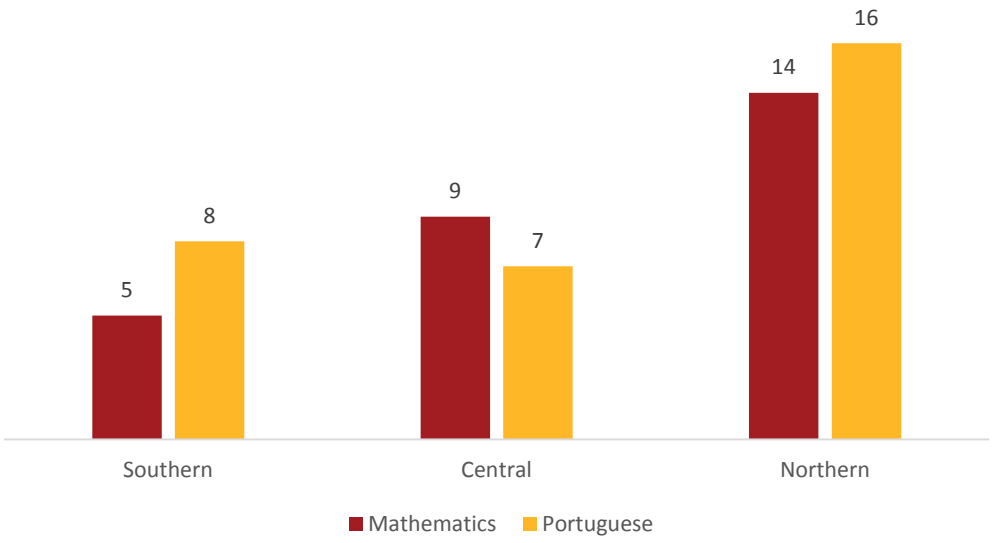
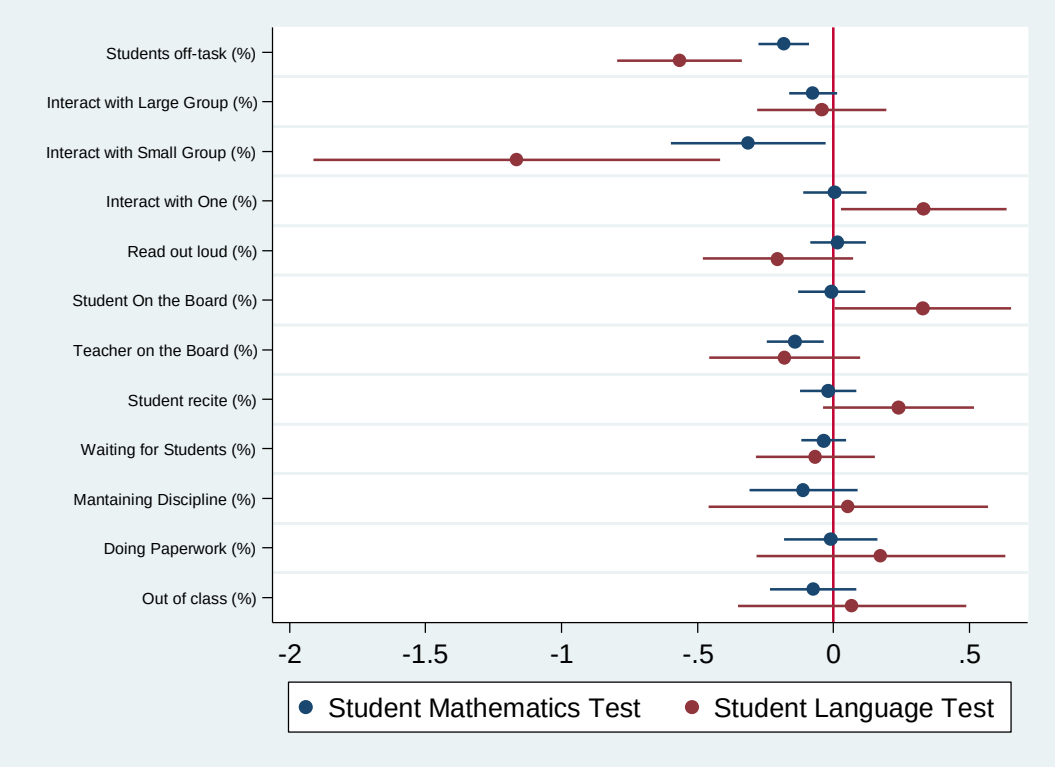


Figure C 3 presents the estimates from a regression of pupils' outcomes (language and mathematics test scores) on teacher use of time after controlling for teacher subject knowledge, pupil non-verbal ability, pupils having breakfast in the morning before the test, as well as province-fixed effects and standard errors clustered at the school level.

While most activities did not provide evidence of a relationship with pupil's outcomes, some did. In particular, for both Portuguese and mathematics classes, interacting with a smaller number of pupils was associated with lower pupils' outcomes. Also, a higher ratio of pupils off task was also correlated with lower outcomes, which was expected.

Figure C 3. Correlates of instructional time and pupil learning



Classroom Practices

This section describes the type of practices adopted by teachers in Mozambique. Table C 2 presents the average classroom practices observed by enumerators during the classroom observation disaggregated by teacher gender, experience, and subject.

Table C 2. Classroom practices by teacher's gender, experience, and subject

Classrooms where the following occurred at least once:	All	Teacher female	Teacher male	< 5 years of experience	> 5 years of	Portuguese	Mathematics
Teacher used her manual (% teachers)	32	46	27	32	31	34	30
Percentage of pupils who used textbook (% pupils)	*44*	*48*	*43*	*41*	*49*	*56*	*36*
Teacher used the board (% teachers)	100	100	100	99	100	99	100
Classrooms where pupils used blackboard (% classrooms)	53	57	52	53	53	36	65
Percentage of pupils who used the board (% pupils)	*29*	*26*	*31*	*30*	*28*	*28*	*30*
Percentage of pupils who use pencil (% pupils)	*92*	*90*	*93*	*95*	*89*	*90*	*94*
Teacher used local information (% teachers)	19	12	21	18	21	25	16
Teacher mostly sitting (% teachers)	5	9	4	5	5	4	6
Teacher approached individual pupils (% teachers)	66	66	65	66	65	62	68
Percentage of pupils approached (% pupils)	*45*	*47*	*45*	*43*	*48*	*42*	*48*
Teacher called pupils by name (% teachers)	62	62	62	60	65	60	63
Percentage of pupils called by name (% pupils)	*35*	*32*	*36*	*36*	*34*	*35*	*35*
Teacher smiled at pupils (% teachers)	41	41	42	41	43	40	42
Teacher scolded/hit pupils (% teachers)	2	6	1	2	3	2	2
Teacher asked questions requiring memorization (% teachers)	52	28	59	55	48	54	50
Teacher used tasks to practice what they have learned (% teachers)	67	57	70	67	67	53	77
Teacher applied new material to other topics (% teachers)	31	26	33	27	37	32	31
Teacher asked question requiring creativity (% teachers)	24	24	24	25	22	25	23
Teacher gave positive reinforcement (% teachers)	40	40	40	36	46	43	39
Teacher corrected pupils errors (% teachers)	60	61	59	56	64	54	64
Teacher gave negative reinforcement (% teachers)	45	34	48	40	52	40	48
Teacher introduced lesson (% teachers)	83	93	80	93	69	76	88
Teacher summarized lesson (% teachers)	17	19	17	14	22	11	22
Teacher gave homework (% teachers)	36	30	37	31	41	27	42
Teacher gave back or took home pupil homework (% teachers)	14	23	12	17	11	9	18
Teacher used local language (% teachers)	40	24	45	44	36	39	41

There were several interesting points to note. While the section on availability of inputs showed that almost 80 percent of pupils had the proper textbook, only 44 percent of them actually used it during the classroom (56 percent in Portuguese class versus 36 percent in mathematics). Only 32 percent of teacher used the teacher manual during class, with a higher percentage of women using it (46 percent versus 27 percent) than men. Almost 20 percent of teachers used local information as a means of contextualizing the lesson, especially male teachers (21 percent) and in the Portuguese classes (25 percent).

When looking at the interaction with pupils, more than half (62 percent) of teachers called pupils by name and about four in 10 teachers smiled and played with pupils and offered positive reinforcement (complemented pupils when answering questions correctly). In mathematics classes, pupils practiced more often what they learned in the lesson than in the language classes (77 percent versus 53 percent), though in only one third of all classes were pupils able to apply those concept to new topics in both Portuguese and mathematics.

Another interesting fact was that while most teachers introduced the lesson (83 percent), and especially female teachers (93 percent), few of them summarized the lesson at the end (17 percent). Finally, in only 36 percent of classes did the teacher give pupils homework and even fewer times (14 percent) did the teacher return corrected homework or take home pupils' work to correct. The numbers were even lower for Portuguese classes (27 percent and nine percent, respectively).

Table C 3 complements the previous table by disaggregating the classroom practices by teacher knowledge and pupil performance. The table displayed the average classroom practices for the quintile of teachers that scored the highest in mathematics, Portuguese, and pedagogy, as well as the classes where the highest performers (the best 25 percent) attended, as well as classes where all pupils received at least 20 percent on the exam.

Table C 3. Classroom practices by teacher knowledge and pupil performance

Classrooms where the following occurred at least once:	All	25% Best Math Teachers	25% Best Language Teachers	25% Best Pedagogy Teacher	Best Performers (Using Average score)	Best Performers (All pupils above 20%)
Teacher used her manual (% of teachers)	32	43	39	37	49	53
Percentage of pupils who used textbook (% pupils)	*44*	*47*	*76*	*53*	*60*	*42*
Teacher used the board (% teachers)	100	100	10	100	100	100
Pupil used board (% pupils)	53	74	4	58	46	73
Pupils who used the board (% pupils)	*29*	*30*	*46*	*27*	*26*	*41*
Pupils who used pencil (% pupils)	*92*	*91*	*93*	*95*	*92*	*100*
Teacher used local information (% teachers)	19	22	16	21	14	27
Teacher mostly sitting (% teachers)	5	0	4	1	0	0
Teacher approached individual pupils (% teachers)	66	84	48	68	69	73
Percentage of pupils approached (% pupils)	*45*	*51*	*52*	*45*	*45*	*62*
Teacher called pupils by name (% teachers)	62	71	65	62	49	73
Percentage of pupils called by name (% pupils)	*35*	*38*	*36*	*30*	*34*	*50*
Teacher smiled at pupils (% teachers)	41	51	44	53	42	50
Teacher scolded/hit pupils (% teachers)	2	2	4	0	8	16
Teacher asked questions that requiring memorization (% teachers)	52	59	42	58	58	42
Teacher used tasks to practice what they have learned (% teachers)	67	83	45	74	72	73
Teacher applied new material to other topics (% teachers)	31	40	13	31	34	34
Teacher asked question requiring creativity (% teachers)	24	26	13	21	22	7
Teacher gave positive reinforcement (% teachers)	40	35	59	47	46	34
Teacher corrected pupils errors (% teachers)	60	62	65	55	63	51
Teacher gave negative reinforcement (% teachers)	45	54	25	44	40	31
Teacher introduced lesson (% teachers)	83	93	96	82	98	100
Teacher summarized lesson (% teachers)	17	31	7	19	33	50
Teacher gave homework (% teachers)	36	62	18	45	45	50
Teacher gave back or took home pupil homework (% teachers)	14	19	5	13	23	58
Teacher used local language (% teachers)	40	39	38	34	32	53

In classes where all pupils received at least 20 percent on the exam, all teachers introduced the lesson and half of them summarized it, which was more than double the occurrence of these practices in the average class. Also, pupils were given more homework (50 percent) and teachers corrected more homework (58 percent) in these classes.

Finally, Table C 4 shows the average classroom practices for other SDI countries. Again, countries where pupils scored higher, like Kenya, there were several classroom practices that were different than classrooms in Mozambique. Teachers gave more homework (71 percent) and corrected homework more often (42 percent); introduced the lesson (91 percent) and summarized the lesson (63 percent) more often; used the teacher manual (83 percent) more than pupils used their

textbooks; used positive reinforcement more frequently; and asked pupils to do tasks to practice what was learned in the classroom, etc.

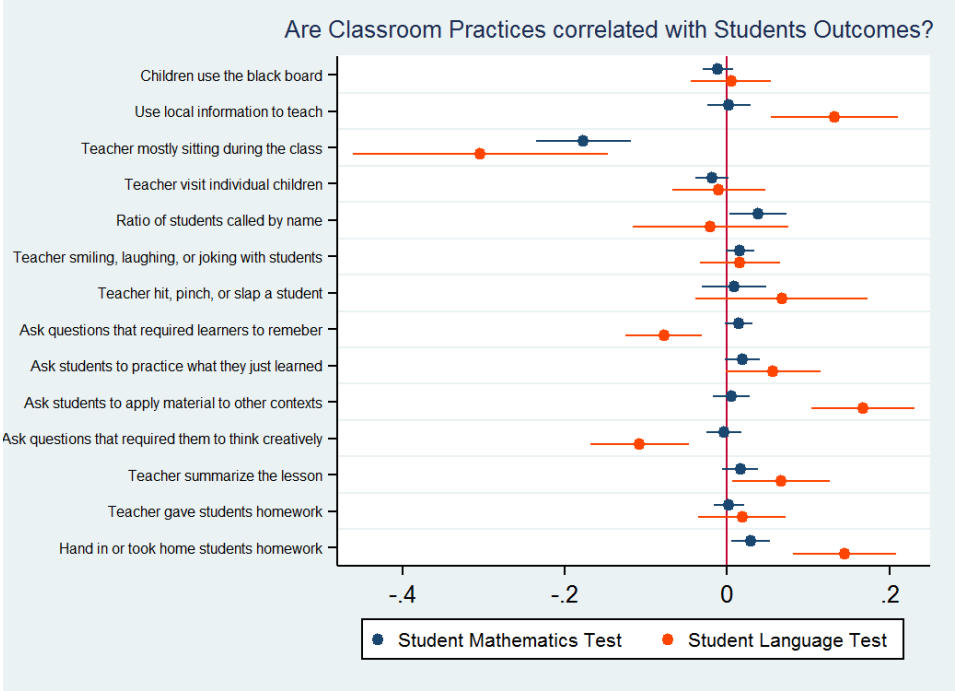
Table C 4. Classroom practices across SDI countries

Classrooms where the following occurred at least once:	Kenya	Nigeria	Tanzania	Togo	Uganda
Teacher used manual (% teachers)	87	64	86	49	72
Number of pupils who used textbook (% pupils)	25	10	14	5	5
Percentage of pupils who used textbook (% pupils)	*78*	*45*	*41*	*15*	*11*
Teacher used blackboard (%teachers)	96	95	100	98	100
Pupils used blackboard (% pupils)	34	27	48	67	39
Number of pupils that used blackboard (% pupils)	5	4	4	4	3
Pupils that used blackboard (% pupils)	*19*	*18*	*15*	*19*	*11*
Number of pupils with pencils (% pupils)	31	21	37	5	41
Pupils who used pencils (% pupils)	*98*	*85*	*95*	*19*	*96*
Teacher used local information (% teachers)	54	59	21	44	28
Teacher mostly sitting (% teachers)	0	9	5	14	7
Teacher approached individual pupils (% teachers)	76	59	58	50	76
Pupils approached (% pupils)	*43*	*27*	*47*	*23*	*33*
Teacher called on pupils by name(% teachers)	87	65	82	96	89
Number of pupils called on by name (% pupils)	8	5	6	11	9
Pupils called on by name (% pupils)	*35*	*28*	*18*	*43*	*25*
Teacher smiled at pupils (% teachers)	75	71	67	50	59
Teacher scolded/hit pupils (% teachers)	4	19	3	10	8
Teacher asked questions requiring memorization (% teachers)	86	77	54	78	88
Teacher used tasks to practice what they learned (% teachers)	75	73	88	73	88
Teacher applied new material to other topics (% teachers)	50	58	68	45	63
Teacher asked questions requiring creativity (% teachers)	57	57	61	44	52
Teacher gave positive reinforcement (% teachers)	87	67	77	82	85
Teacher corrected pupils errors (% teachers)	87	72	85	44	88
Teacher gave negative reinforcement (% teachers)	25	44	29	15	18
Teacher introduced lessons (% teachers)	91	81	88	71	92
Teacher summarized lessons (% teachers)	63	58	48	35	34
Teacher gave homework (% teachers)	71	39	43	44	24
Teacher gave back or collected homework (% teachers)	42	26	20	27	25
Teacher used local language (% teachers)	45	66	2	12	57
Observations	306	760	400	200	400

Figure C 4 presents the estimates from a regression of pupils' outcomes (language and mathematics test score) on classroom practices after controlling for teacher subject knowledge, pupil non-verbal ability, pupils having breakfast in the morning before the test, as well as province-fixed effects and standard errors clustered at the school level.

While most classroom practices did not provide evidence of a relationship with pupil's outcomes, some did. In particular, for Portuguese, having an active teacher who was standing most of the time, who used local information to frame the lesson, who made pupils practice what they had just learned and apply it to a new context, who summarized the lesson at the end of the class, and who regularly corrected pupils homework was correlated with higher pupil scores on the Portuguese test. In mathematics, there were similar results: having an active teacher who called pupils by name, who made pupils practice what they learned, and who regularly gave and corrected homework, led to higher pupil scores on mathematics.

Figure C 4. Classroom practices and pupil learning



For more detail analysis of the relationship between pupils' outcomes and classroom practices, see section *Special Focus: Assessing the relationship of what teachers do, what teachers know, and pupil performance*.

ANNEX D. SPECIAL FOCUS: ASSESSING THE RELATIONSHIP OF WHAT TEACHERS DO, WHAT TEACHERS KNOW, AND PUPIL PERFORMANCE[21]

While it is known that teachers can have a large impact on pupil achievement, at least in the United States, a common finding is that common variables such as experience, education levels, and training of teachers are only weakly related to pupil achievement, a fact documented in Hanushek & Rivkin (2006) and Hanushek & Woessmann (2007). Hanushek & Rivkin (2006) provides an overview of the evidence of the impacts of teacher experience and credentials on pupil achievement. The authors show that among 34 high quality teacher value-added model studies, none showed statistically positive effects of teacher education on pupil achievement and nine percent showed negative effects. For experience, the authors found that only 41 percent showed a positive effect and three percent showed a negative effect.

One characteristic that does seem to predict better pupil performance is the teacher's knowledge level of the subject they are teaching. Metzler & Woessmann (2012), in the context of Peru, found that one standard deviation increase in subject specific teacher knowledge raises pupil achievement by .09 standard deviations in mathematics and has a null effect in reading. Shepherd (2013) examined teacher subject knowledge in South Africa and found that teacher knowledge improves pupil achievement in the wealthiest quintile of schools.

Also, there is some evidence that teacher pedagogical practices and instructional time affect pupil achievement. Lavy (2010) examines the impact of the length of instructional time in explaining pupil achievement using PISA 2006 and Israeli administrative data and finds that instructional time has a positive and statistically significant impact on pupil achievement. The estimates suggest an additional hour of instruction per week increases test scores by .07 standard deviations. Lavy (2011) examines pedagogical practices of teachers and finds that the practice of "instillment of knowledge and comprehension" and endowing pupils with analytical and critical skills have large impacts relative to other teaching practices. Dobbie & Fryer, Jr. (2013) find increased instructional time and high expectations predict school effectiveness in a sample of charter schools in New York City. A 25 percent increase in instructional time was associated with an increase in mathematics achievement of .05 standard deviations. They also find that schools with high academic and behavioral expectations have mathematics scores that are .044 standard deviations higher and reading scores, .03 standard deviations higher.

[21] This section borrowed heavily from Filmer, Molina, and Stacy (2015).

Filmer et al (2015) examine the roles of human capital, instructional time, and particular teacher practices in raising pupil achievement. These three dimensions of teacher quality capture the following: do the teachers have the skills to teach (human capital), do they spend time in the classroom (effective instructional time), and what do the teachers do in the classroom. We can write the teacher effect as a function of the three dimensions and unobservables η_j.

$$\tau_j = F(H_j, I_j, P_j) + \eta_j, \quad (1)$$

where H_j is a vector of human capital characteristics of the teacher, I_j is instructional time, and P_j are the instructional practices of the teacher. We model (1) as a linear function of the characteristics, but allow for an interaction between human capital and instructional time to test whether additional instructional time is particularly effective after a teacher reaches a threshold in terms of human capital.

$$\tau_j = H_j\beta_1 + I_j\beta_2 + I_j H_j^* \beta_3 + P_j\beta_4 + \eta_j, \quad (2)$$

where H_j^* denotes a teacher with high levels of human capital.[22]

We use the subject knowledge examinations and the teacher pedagogical examinations as measures of teacher human capital. As discussed previously, the pedagogical tests examined a teacher's ability to prepare a lesson, assess pupil abilities, and evaluate pupil progress in learning.[23] We form a measure of effective instructional time by combining data from the staff roster module (used to measure absence rate), the classroom observation module, and reported teaching hours for the school day. The instructional time for the school day is adjusted for the time teachers are absent from the classroom, on average, and for the time the teacher teaches while in classrooms based on classroom observations recorded every one minute in a teaching lesson. We examine three particular class room practices: the teacher instilling trust in the pupils, the teacher instilling discipline in the pupils, and the teacher challenging pupils intellectually.[24]

[22] For our purposes, this means that the teacher answered more than 50 percent of the subject knowledge questions correctly and more than 25 percent of the pedagogy questions correctly. These levels of subject knowledge and pedagogical skill are roughly at the average for the sample of teachers.

[23] The teachers are asked to prepare a lesson based on a news article presented to them. Then they are asked to read two paragraphs from two hypothetical pupils and judge the strengths and weakness. Finally, they are asked to evaluate a set of test scores for pupils.

[24] The "instill trust" measure is an indicator for whether the teacher called on pupils by name and praised pupils while giving feedback. The "instill discipline" measure is an indicator for whether the teacher hit or slapped pupils during the lesson and scolded while giving feedback during a lesson. The "challenge pupil intellectually" measure is an indicator for whether the teacher assigned or reviewed homework in class and asked questions of pupils during the lesson.

Table D 1 presents the cross-sectional results acknowledging that they are not causal. The first two columns of the table shows the effects of human capital alone on pupil achievement. In the middle two columns, human capital and instructional time are included as covariates. Additionally, in order to test for a particularly strong effect for instructional time if it is paired with high human capital, instructional time is interacted with an indicator for whether the teacher has a high level of human capital, which is defined as whether the teacher answered more than 50 percent of the teacher subject knowledge questions correctly and more than 25 percent of the pedagogy questions correctly. These levels of subject knowledge and pedagogical skill are roughly at the average for the sample of teachers. Finally, in columns 5 and 6, particular teacher practices are also included as covariates.

Table D 1. Estimated effects of teacher human capital, instructional time, and practices on pupil achievement

Variables	Human Capital		Human Capital & Instruction Time		Human Capital, Instruction Time, & Practices	
	Math	Reading	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher knowledge	0.087***	0.033**	0.085***	0.030	0.088***	0.028
	(0.018)	(0.016)	(0.021)	(0.021)	(0.021)	(0.021)
Pedagogical skill	0.046***	0.101***	0.054**	0.105***	0.051**	0.104***
	(0.015)	(0.015)	(0.022)	(0.022)	(0.022)	(0.022)
Effective instructional time (hours)			0.034***	0.054***	0.034***	0.053***
			(0.011)	(0.011)	(0.011)	(0.011)
Instructional time-high human			-0.013	-0.018	-0.014	-0.018
			(0.012)	(0.011)	(0.012)	(0.011)
Instill trust					0.073**	0.069*
					(0.037)	(0.040)
Instill discipline					-0.138	-0.087
					(0.092)	(0.076)
Challenge pupils intellectually					0.093**	0.159***
					(0.041)	(0.040)
District Fixed Effects	X	X	X	X	X	X
Classroom Random Effects	X	X	X	X	X	X
Pupil covariates	X	X	X	X	X	X
Teacher covariates	X	X	X	X	X	X
School covariates	X	X	X	X	X	X
Observations	11274	12216	8284	8574	8080	8564
Number of teachers	1342	1424	959	967	933	966

Notes: Effects estimated using a classroom random effects regression. Pupil covariate set includes: pupil non-verbal reasoning ability, age, gender, whether the pupil ate breakfast, and the native language of the pupil. Teacher covariate set includes: teacher gender, experience, university degree status, education degree status, head teacher status, and contract teacher status. School covariates set includes: whether the school is in an urban location, whether it is a private school, the school infrastructure index, equipment index, pupil-teacher ratio, the school non-verbal reasoning level, and whether the school has a PTA or governing board. High human capital is defined as teacher answering more than 50% of the subject knowledge questions correctly and 25% of the pedagogy questions.

Both forms of human capital, teacher knowledge, and pedagogical skill, had a statistically significant impact on pupil achievement. The estimates suggest that teacher subject knowledge was an important predictor of pupil achievement, even after conditioning on pedagogical skill. The coefficient in mathematics was .087 in mathematics and .033 in reading in columns 1 and 2. The coefficient on teacher pedagogical skill for mathematics was .046 and statistically significant at the one percent level, meaning a one standard deviation increase in pedagogical skill increased pupils' achievement in mathematics by .046 standard deviations in the test score. The coefficient for reading was .101 and significant at the one percent level. The estimates for teacher knowledge and pedagogical skill were similar in columns 3 through 6 after controlling for instructional time and practices.

Teacher instructional time also had a statistically significant positive relationship for mathematics and reading. The estimated coefficient for mathematics, shown in column 3, is .034, suggesting that an additional hour of effective instructional time increased pupil achievement by .034 standard deviations of the test score. The estimated coefficient on reading was .054. Interestingly, the estimated interaction effects with human capital were statistically insignificant, meaning there was little evidence that instructional time had a particularly strong effect for teachers with high human capital. Lavy (2010) reports an effect size of .07 standard deviations for an additional hour of instructional time per week in the context of PISA countries. This implies the effect of an additional hour per day is roughly .35 in his data, which is much larger than the effect estimated in the SDI data. Dobbie & Fryer, Jr. (2013) reports an effect of .05 standard deviations in mathematics for an increase of 25 percent in instructional time over the average number of hours in public schools. Assuming the average number of hours per day in public schools is around 6.5, which indicates an increase of around 1.625 hours, this implies an effect per additional hour of around .03 standard deviations, which is similar to what we found.

Of the specific classroom practices, which were all binary indicators of whether or not the teacher engaged in this practice during their classroom observation, the practice of instilling trust had a positive and statistically significant impact on mathematics and reading. The estimated coefficient for mathematics was .073 and the estimate for reading was .069. The effect for the practice of instilling discipline had a negative point estimate for both mathematics and reading, but was insignificant. The effect of challenging pupils intellectually by assigning homework and asking questions was statistically significant for mathematics and reading. The estimated effect for mathematics is .093 and the estimate for reading is .159. For comparison, Dobbie & Fryer, Jr. (2013) found that charter schools with high academic and behavioral expectations had .044 standard deviation higher mathematics and .03 standard deviation higher reading achievement. We view this measure as broadly similar to our measure of challenging pupils intellectually.

Table D 2 replicates this exercise only for the Mozambique data (rather than all SDI available countries). Before interpreting the results, a note of caution is necessary. Since, in order to replicate this exercise, we needed teachers that had been tested for knowledge, whose pupils took the pupil test, and whose class had been observed, the numbers of observations available diminished considerable. When we add the earlier concerns about the identification strategy, findings should be interpreted with extreme care. The results point to the fact that teachers' time with pupils could only have a transformative effect on pupils' performance if teachers had a minimum level of competence. It was only for these teachers that additional time teaching was useful in improving pupil performance. Also, positive reinforcement practices were correlated with better pupil outcomes in both mathematics and Portuguese.

Table D 2. Estimated effects of teacher human capital, instructional time, and practices on pupil achievement (Mozambique sample)

Variables	Human Capital		Human Capital & Instruction Time		Human Capital, Instruction Time, & Practices	
	Math	**Reading**	**Math**	**Reading**	**Math**	**Reading**
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher knowledge	0.031	0.123*	0.026	0,098	0.036*	0.092
	(0.023)	(0.074)	(0.024)	(0.074)	(0.022)	(0.077)
Pedagogical skill	-0/007	-0.037	-0.037	-0.056	-0.026	-0.052
	(0.026)	(0.049)	(0.027)	(0.054)	(0.025)	(0.053)
Effective instructional time (hours)			0.006	0.029	0.005	0.032
			(0.018)	(0.026)	(0.017)	(0.025)
Instructional time-high human capital			0.066***	0.853***	0.054**	0.870***
			(0.024)	(0.168)	(0.025)	(0.185)
Instill trust					0.119***	0.138*
					(0.039)	(0.080)
Instill discipline					0.048	0.095
					(0.147)	(0.308)
Challenge pupils intellectually					-0.017	0.143
					(0.050)	(0.096)
District Fixed Effects	X	X	X	X	X	X
Classroom Random Effects	X	X	X	X	X	X
Pupil covariates	X	X	X	X	X	X
Teacher covariates	X	X	X	X	X	X
School covariates	X	X	X	X	X	X
Observations	1451	1462	1399	1410	1389	1400
Number of teachers	167	168	160	161	159	160

Notes: Effects estimated using a classroom random effects regression. Pupil covariate set includes: pupil non-verbal reasoning ability, age, gender, whether the pupil ate breakfast, and the native language of the pupil. Teacher covariate set includes: teacher gender, experience, university degree status, education degree status, head teacher status, and contract teacher status. School covariates set includes: whether the school is in an urban location, whether it is a private school, the school infrastructure index, equipment index, pupil-teacher ratio, the school non-verbal reasoning level, and whether the school has a PTA or governing board. High human capital is defined as teacher answering more than 50% of the subject knowledge questions correctly and 25% of the pedagogy questions.

ANNEX E. USE OF LOCAL LANGUAGE, TEMPORARY CLASSROOMS AND PUPIL LEARNING

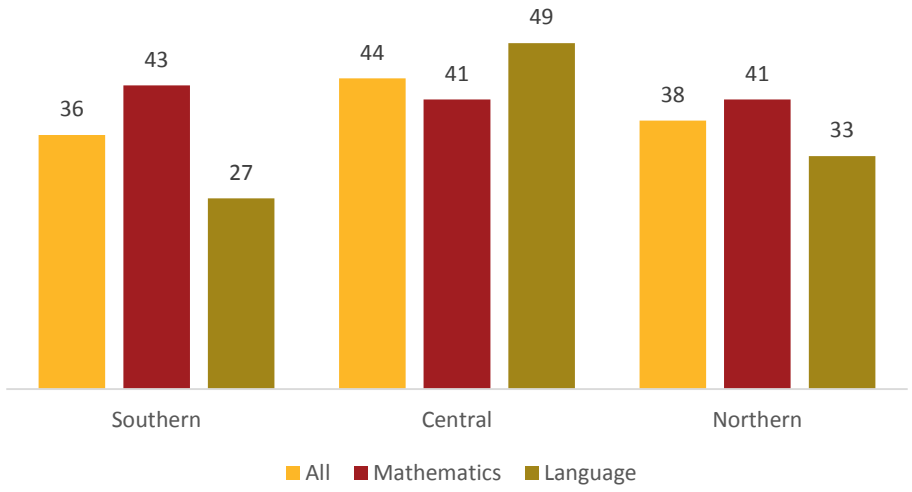
This section will attempt to answer two questions: (i) Does the use of local language as a means of instruction improve pupil learning? and (ii) Does having a class under a tree as compared to a proper classroom diminish pupil learning?

We acknowledge that we do not have the proper identification strategy to claim we can answer these questions, but below we propose an exploratory analysis with the data we collected.

Local Language and Pupil Learning

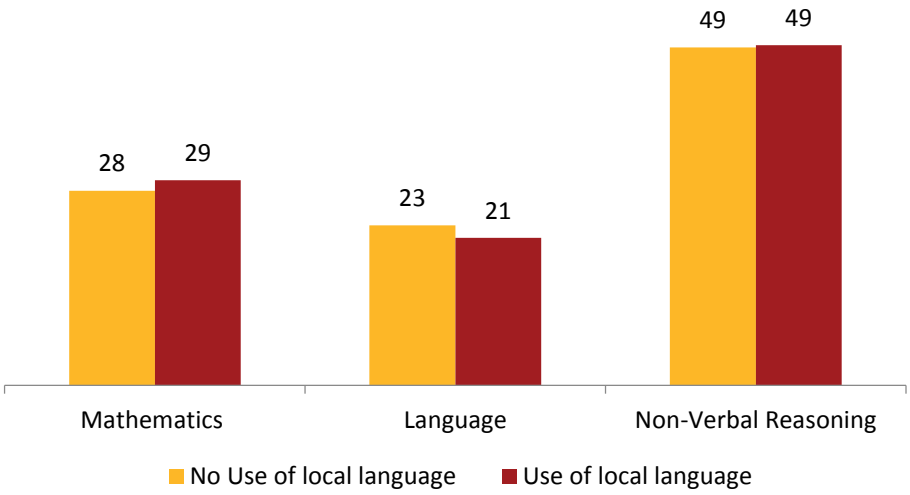
Currently more than a third of fourth grade teachers use local language as a mean of instruction in both Portuguese and mathematics classes (Figure E 1).

Figure E 1. Use of local language as a mean of instruction by region (percent)



Pupils whose teachers use local language during class have statistically equal scores to pupils whose teachers do not use local language as a means of instruction. Figure E 2 presents the results.

Figure E 2. Test scores and use of local language as a mean of instruction (percent)



Of course, using local language could be related to unobservable characteristics of the pupils, as well as innate ability. To understand whether similar pupils receiving similar classes, one where the teacher uses local language and one where the teacher does not, we need to do further analysis. Below we present a series of regressions to study this relationship. All regressions include province fixed effects, as well as cluster standard errors at the school level. We use ordinary least squares (OLS) to run all regressions, but in some we use coarsened exact matching (CEM) to select a sample of similar teachers and pupils across which we could run the regression. In particular, we match on teacher subject knowledge, pupil non-verbal ability, and whether pupils had breakfast on the day of the exam. Because of this, the sample for those regressions is substantially smaller.

Table E 1 present the results. The evidence from the six different regressions does not support the theory of a positive relationship between those two variables. If anything, the evidence points to a zero or small negative effect of using local language on pupil learning. As we acknowledged before, our weak identification strategy does not allow us to give any definite answer and more research would be needed to know the causal effect of using local language on pupil learning.

Table E 1. Relationship between test scores and use of local languages

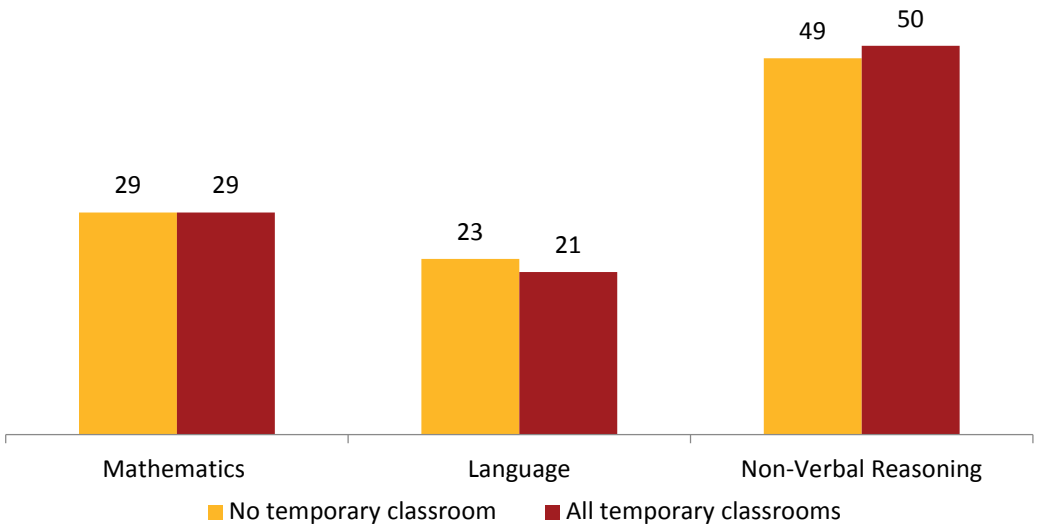
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS Mathematics	OLS Mathematics	CEM Mathematics	OLS Language	OLS Language	CEM Language
Local language	-0.0279**	-0.0113	-0.0612	-0.0793**	-0.0408	-0.182*
	(0.0109)	(0.0116)	(0.0360)	(0.0307)	(0.0385)	(0.0976)
Teacher mathematics		0.0477**	0.102			
		(0.0189)	(0.217)			
Teacher pedagogy score		0.0101	-0.0216		-0.00109	-0.305
		(0.0305)	(0.202)		(0.125)	(0.863)
Pupils non-verbal reasoning		0.125***	0.164*		0.230***	0.353***
		(0.0131)	(0.0833)		(0.0358)	(0.119)
Breakfast		0.000323	0.00525		0.0193	-0.148
		(0.00887)	(0.0655)		(0.0206)	(0.0890)
Teacher language score					0.236	0.723
					(0.158)	(1.192)
Constant	0.266***	0.168***	0.187	0.258***	-0.0462	0.147
	(0.00565)	(0.0161)	(0.113)	(0.0199)	(0.0793)	(0.379)
Observations	1,545	1,418	83	1,545	1,429	195
R2	0.012	0.142	0.176	0.015	0.123	0.163

Note: Robust standard errors in parentheses. Levels of significance: *** p<0.01, ** p<0.05, * p<0.1

ANNEX F. CLASSROOM UNDER A TREE AND PUPIL LEARNING

Pupils taught in temporary classrooms (classroom under a tree and other temporary classrooms made of mud) have statistically equal scores to pupils taught in proper classrooms. Figure F 1 presents the results.

Figure F 1. Temporary classrooms and pupil learning (percent)



Of course, again, temporary classrooms could be related to unobservable characteristics that prevent us from inferring causality from the correlation of these two variables. To understand whether similar pupils receiving similar classes, ones where they are taught in temporary classrooms and the others in proper classrooms, we present a series of regressions below. All regressions include province fixed effects as well as cluster standard errors at the school level. We use ordinary least squares (OLS) to run all regressions, but in some regressions we use coarsened exact matching (CEM) to select a sample of similar teacher and pupils across which we could run the regression. In particular, we matched on teacher subject knowledge, pupil non-verbal ability, and whether pupils had breakfast on the day of the exam. Because of this, the sample for those regressions is substantially smaller.

Table F 1 present the results. The evidence from the six different regressions does not support the theory of a negative relationship between those two variables. If anything, the evidence point to a no effect of temporary classrooms on pupil learning. As we acknowledged before, our weak identification strategy does not allow us to give any definite answer and more research would be needed to identify the causal effect of temporary classrooms on pupil learning.

Table F 1. Relationship between test scores and temporary classrooms

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS Mathematics	OLS Mathematics	CEM Mathematics	OLS Language	OLS Language	CEM Language
Teacher mathematics score		0.0494**	0.0781*			
		(0.0202)	(0.0462)			
Teacher pedagogy score		0.0110	-0.00340		0.0104	0.221
		(0.0301)	(0.107)		(0.127)	(0.263)
Pupils non-verbal reasoning		0.126***	0.153***		0.234***	0.153**
		(0.0132)	(0.0275)		(0.0354)	(0.0720)
Breakfast		0.00254	0.00851		0.0215	0.133***
		(0.00875)	(0.0195)		(0.0214)	(0.0341)
Temporary classrooms	-0.00155	0.0189	0.0223	-0.00211	-0.0015	0.00493
	(0.0133)	(0.0133)	(0.0238)	(0.0310)	(0.0385)	(0.0462)
Teacher language score					0.270*	0.401
					(0.150)	(0.403)
Constant	0.254***	0.157***	0.154***	0.219***	-0.0859	-0.0980
	(0.00519)	(0.0158)	(0.0273)	(0.0165)	(0.0664)	(0.136)
Observations	1,761	1,418	351	1,761	1,429	706
R2	0.000	0.143	0.139	0.000	0.120	0.076

Note: Robust standard errors in parentheses. Levels of significance: *** p<0.01, ** p<0.05, * p<0.1

ANNEX G. ADDITIONAL RESULTS

Table G 1. Correlates of teacher effort

VARIABLES	(1) Absent from school or class	(2) Absent from school or class	(3) Absent from school or class	(4) Absent from school or class
Female	0.0421			0.0294
	(0.0378)			(0.0379)
Have a training certificate	-0.0328			-0.00952
	(0.0629)			(0.0613)
Born in the district	-0.0211			-0.0212
	(0.0405)			(0.0414)
Experience (years)	-0.00272			-0.00128
	(0.00283)			(0.00279)
Secondary education	-0.0262			0.0264
	(0.0764)			(0.0795)
University degree	-0.0869			-0.0184
	(0.104)			(0.109)
Contract teacher	0.00308			-0.0189
	(0.0458)			(0.0458)
Urban	-0.0971	-0.0667		-0.0104
	(0.0615)	(0.0586)		(0.0658)
School has a paved road	-0.152*	-0.158**		-0.159**
	(0.081)	(0.0737)		(0.0724)
Infrastructure index		-0.0439		-0.0209
		(0.0488)		(0.0511)
Teacher-pupil ratio		-0.00519***		-0.00492**
		(0.0018)		(0.00225)
Equipment index		-0.0863		-0.0887
		(0.0561)		(0.0617)
Electricity		0.0641		0.031
		(0.0756)		(0.0876)
Share of pupils with textbook		-0.127*		-0.163*
		(0.0742)		(0.0853)
Director(s) absence (%)			0.252***	0.209***
			(0.059)	(0.0596)
Inspected in the last 30 days			-0.0339	-0.0386
			(0.051)	(0.052)
Distance to government office more than 30 minutes			0.0517	0.0358
			(0.0505)	(0.0473)
Delays with payments			-0.0742**	-0.0664*
			(0.0367)	(0.0375)
Constant	0.740***	0.962***	0.575***	0.873***
	(0.184)	(0.144)	(0.146)	(0.203)
Observations	812	1006	761	755
R2	0.124	0.124	0.155	0.18

Note: Robust standard errors in parentheses. Levels of significance: *** p<0.01, ** p<0.05, * p<0.1. This includes district fixed effects as well as controls for the day of the week the unannounced visit took place.

Table G 2. Teacher evaluation breakdown

(Out of 100)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Portuguese and mathematics	**34.6**	**33.8**	**34.8**	**0.0**	**39.4**	**33.2**	**33.8**
Portuguese (complete test)	**32.3**	**34.6**	**31.7**	**0.1**	**36.4**	**33.0**	**27.9**
Grammar	79.0	83.2	78.1	0.1	85.5	78.4	75.6
Cloze task	31.5	33.0	31.1	0.1	35.0	35.7	20.4
Composition	9.3	11.2	8.9	0.3	12.7	8.7	8.2
Mathematics (complete test)	**30.4**	**26.7**	**31.2**	**-0.1**	**36.2**	**27.3**	**32.6**
Lower primary	36.8	32.8	37.7	-0.1	42.7	33.2	40.0
Upper primary	18.5	15.2	19.2	-0.2	24.1	16.3	18.6
Comparing fractions	16.8	11.4	18.0	-0.4	16.3	20.1	10.4
Interpreting a Venn diagrams	20.3	22.0	19.9	0.1	28.3	17.3	20.5
Interpreting data on a graph	12.5	8.1	13.5	-0.4	16.3	11.1	12.8
Pedagogy (complete test)	**12.3**	**7.6**	**13.3**	**-0.4*** **	**16.5**	**10.8**	**12.1**
Lesson preparation	16.7	13.0	17.5	-0.3*	19.8	16.3	15.2
Pupil comparisons	10.4	4.3	11.8	-0.6***	16.0	7.3	12.5
Pupil evaluations	5.3	1.7	6.1	-0.7***	9.0	5.1	3.1

Notes: Weighted means using sampling weight and the sample design. Results for teacher knowledge based on observations from 673 sampled teachers. Levels of significance: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table G 3. Pupil assessment: Boys

(Out of 100)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Overall test score	**22.8**	**24.3**	**22.7**	**0.1**	**34.4**	**23.6**	**17.1**
Literacy score	**20.8**	**22.1**	**20.6**	**0.1**	**34.7**	**21.6**	**14.0**
Identify a letter	40.9	36.0	41.4	-0.1	55.5	46.0	26.7
Identify basic words	22.8	23.0	22.8	0.0	40.6	23.4	15.8
Read sentence	15.2	16.9	15.0	0.1	26.3	15.9	9.7
Read paragraph	8.1	10.4	7.8	0.3	15.8	9.2	3.1
Comprehension score	5.9	8.9	5.5	0.6	10.1	7.1	2.2
Numeracy score	**27.1**	**27.3**	**27.1**	**0.0**	**29.0**	**27.2**	**26.2**
Single-digit addition	54.2	48.9	54.8	-0.1	59.7	54.4	51.5
Double-digit addition	23.3	26.4	22.9	0.1	24.1	24.7	20.6
Single-digit subtraction	34.0	39.0	33.4	0.2	38.6	36.6	28.0
Double-digit subtraction	5.6	5.5	5.6	0.0	7.1	5.0	6.1
Single-digit multiplication	4.9	2.2	5.2	-0.6*	5.6	5.4	3.7
Double-digit multiplication	0.1	0.3	0.1	1.7	0.2	0.0	0.3
Single-digit division	10.3	6.6	10.7	-0.4	11.4	13.2	5.2
Non-verbal reasoning score	**44.3**	**54.1**	**43.2**	**0.3*** **	**51.2**	**46.4**	**38.2**

Note: Levels of significance: *** p<0.01, ** p<0.05, * p<0.1.

Table G 4. Pupil assessment: Girls

(Average score in percent)	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Overall test score	**18.3**	**19.2**	**18.1**	**0.1**	**39.1**	**14.9**	**11.0**
Literacy score	**16**	**16.7**	**15.9**	**0.1**	**40.8**	**11.7**	**7.7**
Identify a letter	33.3	35.5	32.9	0.1	52.8	34.0	20.9
Identify basic words	17.8	18.0	17.7	0.0	47.2	14.1	5.9
Read sentence	9.6	8.4	9.8	-0.1	34.5	4.5	2.2
Read paragraph	5.1	5.3	5.1	0.0	20.1	2.0	0.8
Comprehension score	4.1	5.9	3.8	0.5	16.1	2.0	0.2
Numeracy score	**22.6**	**23.2**	**22.5**	**0.0**	**28.2**	**22.2**	**20.0**
Single-digit addition	39.9	35.0	40.8	-0.1	50.8	41.2	31.7
Double-digit addition	12.1	18.3	11.0	0.7*	21.8	10.6	8.5
Single-digit subtraction	21.4	20.8	21.5	0.0	37.9	22.3	10.5
Double-digit subtraction	3.4	1.6	3.7	-0.6*	5.7	3.9	1.2
Single-digit multiplication	2.9	1.6	3.1	-0.5	8.0	1.7	1.6
Double-digit multiplication	0.1	0.0	0.1	-1.0	0.0	0.2	0.0
Single-digit division	6.3	4.7	6.6	-0.3	9.3	7.3	3.2
Non-verbal reasoning score	**43.8**	**50.7**	**42.5**	**0.2*** **	**52.4**	**45.5**	**36.2**

Note: Levels of significance: *** p<0.01, ** p<0.05, * p<0.1.

Figure G 1. Geographic differences in pupil learning (percent)

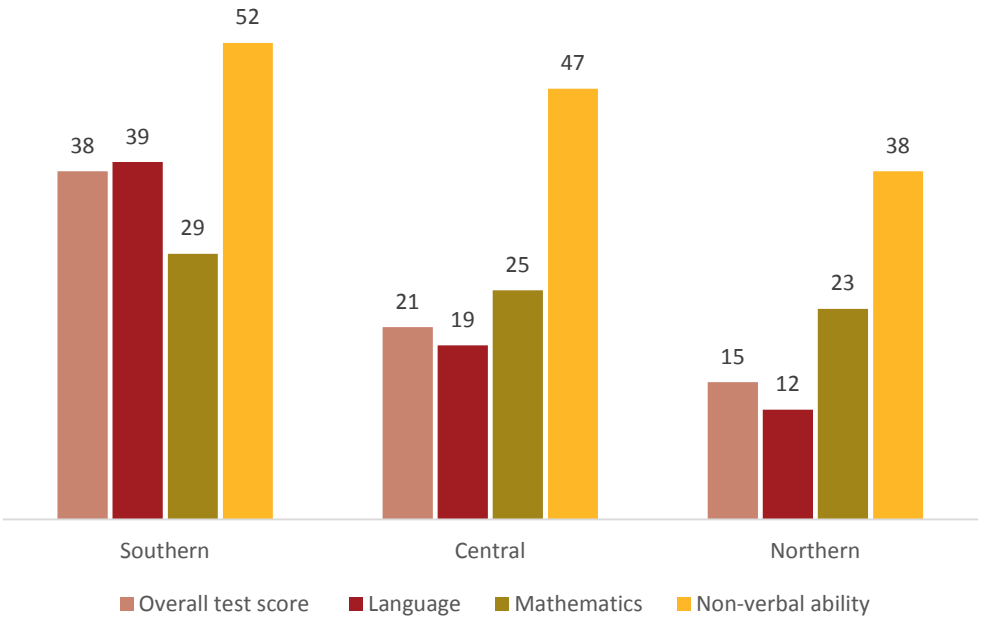


Table G 5. Inequality of outcomes (out of 100)

	Pupil overall test score			Pupil numeracy score			Pupil literacy score			Pupil non-verbal reasoning score		
	No	Yes	Difference	No	Yes	Difference	No	Yes	Difference	No	Yes	Difference
Pupil female	25.76	22.89	2.87*	27.47	23.8	3.67***	24.3	21.5	2.79	46.63	45.66	0.96
School in urban location	24.39	24.41	-0.02	25.66	26.06	-0.39	23.06	22.5	0.56	44.64	54.14	-9.50***
Pupil ate breakfast	22.21	25.14	-2.93*	25.54	25.79	-0.24	20.41	23.85	-3.44*	42.82	47.31	-4.50***
Pupil had protein for breakfast	24.22	24.67	-0.45	25.91	25.43	0.48	22.77	23.28	-0.5	44.81	48.29	-3.48***

Note: Levels of significance: *** p<0.01, ** p<0.05, * p<0.1.

Figure G 2. Pupil evaluation distribution by section

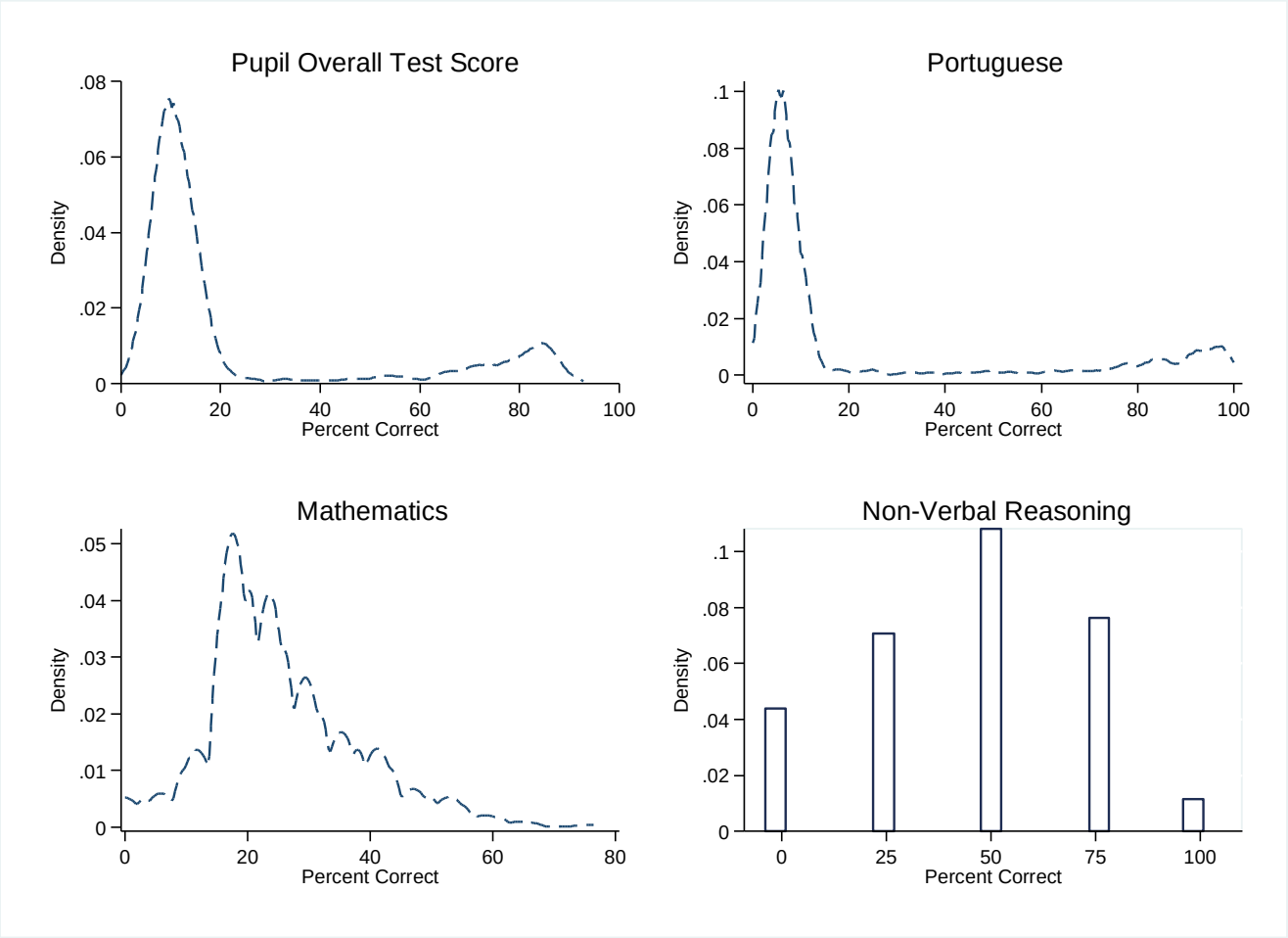
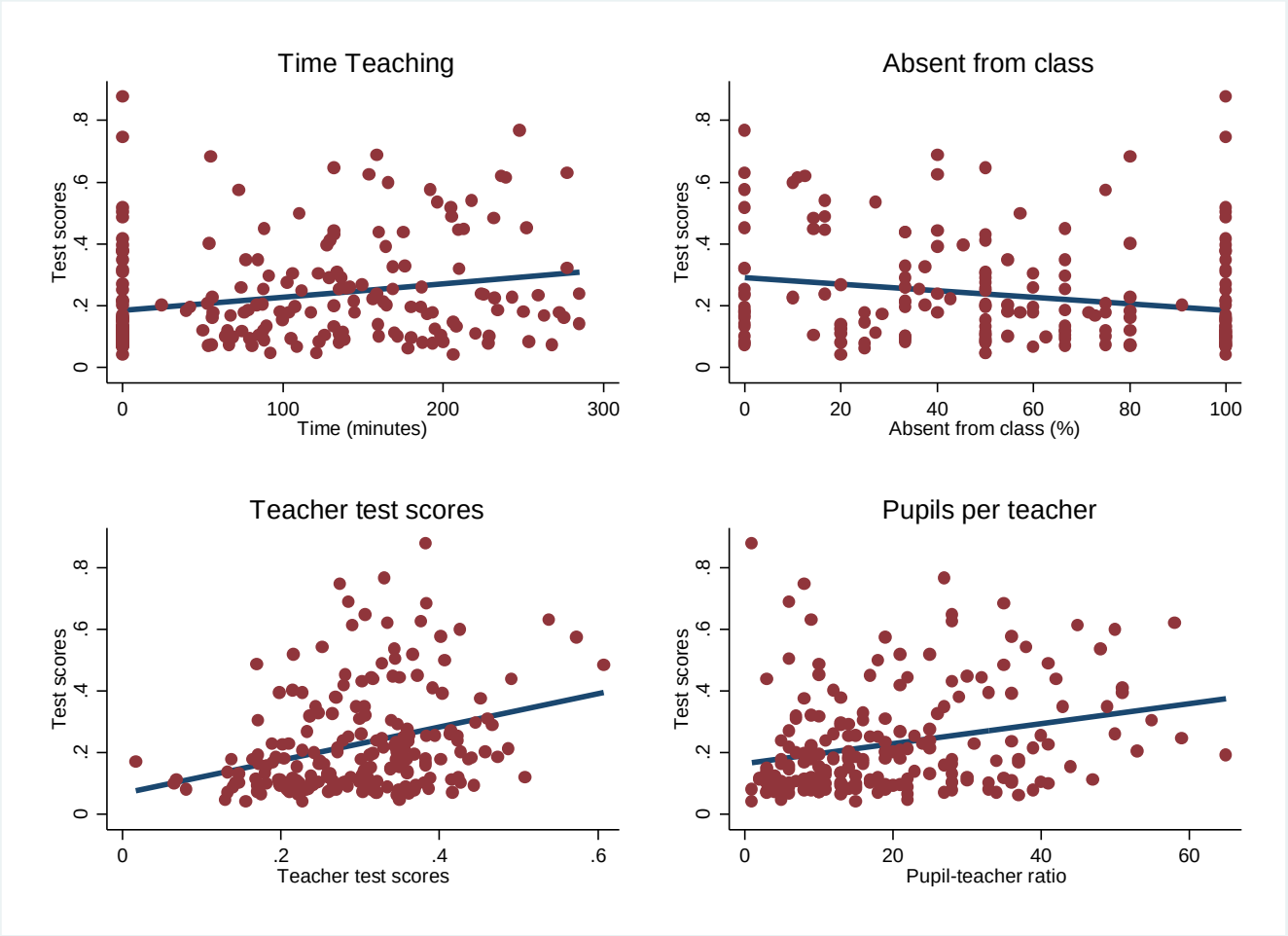


Table G 6. Correlations between the Service Delivery Indicators and pupil test scores

	School absence rate	Absence from class (at school)	Time spent teaching per cay	Share of teachers with minimum knowledge	Teacher test score (language and mathematics)	Teacher test score (language, mathematics, pedagogy)	Minimum equipment availability	Minimum infrastructure availability	Observed Pupil-teacher ratio	Pupils per textbook
Panel A: School Level Results										
All Schools	-0.269	-0.0013	0.0818	0.000174	0.0260***	0.0370***	0.00458***	0.000158	0.0167**	0.00195
	(0.230)	(0.00228)	(0.0539)	(0.00357)	(0.00569)	(0.00719)	(0.00141)	(0.00156)	(0.00652)	(0.00309)
Observations	200	161	200	200	200	200	200	200	200	200
Panel B: Pupil Level Results										
All Schools	-0.388***	0.0249	0.0861***	-0.231	1.884***	2.340***	0.379***	0.132*	0.0123***	0.460***
	(0.0798)	(0.0932)	(0.0201)	(0.29)	(0.247)	(0.299)	(0.0555)	(0.0693)	(0.00211)	(0.0871)
Observations	1,731	1,438	1,731	1,731	1,731	1,731	1,731	1,731	1,731	1,731

Note: Each cell represent a regression where the standardized test score is regressed on the indicator noted in the column and a constant. Time spent teaching is measured in hours. Robust standard errors in parentheses. Data are weighted. Levels of significance: *** p<0.01, ** p<0.05, * p<0.1

Figure G 3. Correlations between indicators and learning (pupil test scores)



Note: The graphs show the scatter plots (dots) and the predicted OLS relationship (solid line) for various indicators and pupil test scores in public schools.

Figure G 4. Absence rates of teachers and directors (percent)

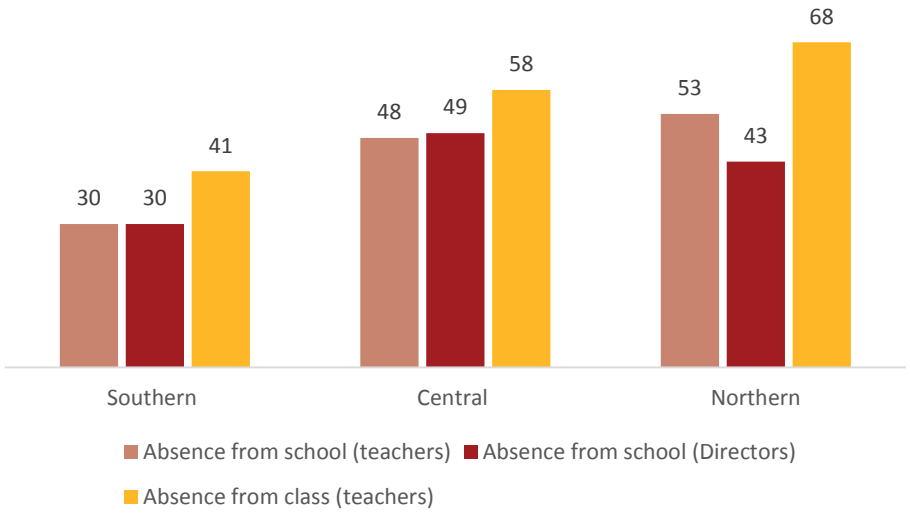


Figure G 5. Teacher scores

Pedagogical skills are consistently below teachers' mathematics and language test scores

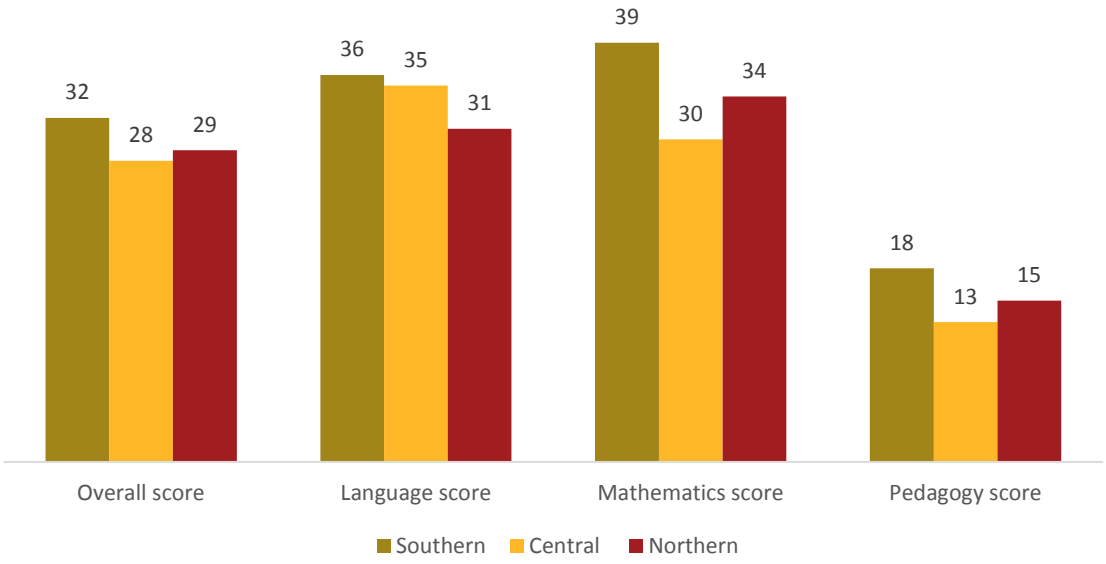


Table G 7. Teacher characteristics (absence rate sample)

	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Age	30.9	35.2	29.5	0.2***	33.1	30.1	30.2
Experience	6.9	10.7	5.7	0.9***	9.5	5.8	6.3
Female	34.2	51.9	29.2	0.8***	47.9	33.8	21.6
Born in district	27.3	28.4	26.9	0.1	35.5	22.2	28.8
Primary	4.7	2.4	5.4	-0.5**	6.9	3.1	5.6
Secondary	85.4	72.0	89.7	-0.2***	75.8	91.7	83.0
Teacher has university degree	9.8	24.9	4.9	4.1***	16.8	5.2	11.4
Has training certification	89.8	88	90.4	0.0	85.1	89.6	95.4
Teacher has bachelor of education degree	1.5	4.0	0.1	25.9***	3.0	1.0	0.7
Director/Head Teacher	8.0	2.4	11.0	-0.8***	6.2	7.8	10.6
Deputy Head Teacher	6.6	3.2	8.5	-0.6***	4.8	7.3	7.5
Government teacher	61.4	74.4	54.4	0.4***	83.8	59.4	38.9
Contract teacher	38.5	25.6	45.4	-0.4***	16.2	40.3	61.1

Notes: Weighted means using sampling weight and the sample design. Results for teachers characteristics from absence rate module and based on observations from 1,006 sampled teachers from 200 schools. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

Table G 8. Teacher characteristics (assessment sample)

	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Teacher age	31.4	34.3	30.7	0.1***	32.8	30.8	31.4
Teacher experience	7.5	9.9	7.0	0.4**	9.3	6.8	7.7
Female	33.6	62.9	27.1	1.3***	52	28.0	31.0
Primary	5.4	12.4	3.9	2.2*	12.3	2.5	6.3
Secondary	88.4	74.7	91.4	-0.2***	70.7	95.4	87.3
University degree	6.0	11.8	4.7	1.5*	17	2.1	5.6
Has training certificate	96.4	99.0	95.8	0.0*	99.1	97.1	92.9
Teacher with education Degree	2.7	7.7	1.6	3.7*	5.8	2.6	0.7

Notes: Weighted means using sampling weight and the sample design. Results for teacher knowledge based on observations from 673 sampled teachers. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

Figure G 6. Distribution of the effort indicators[25]

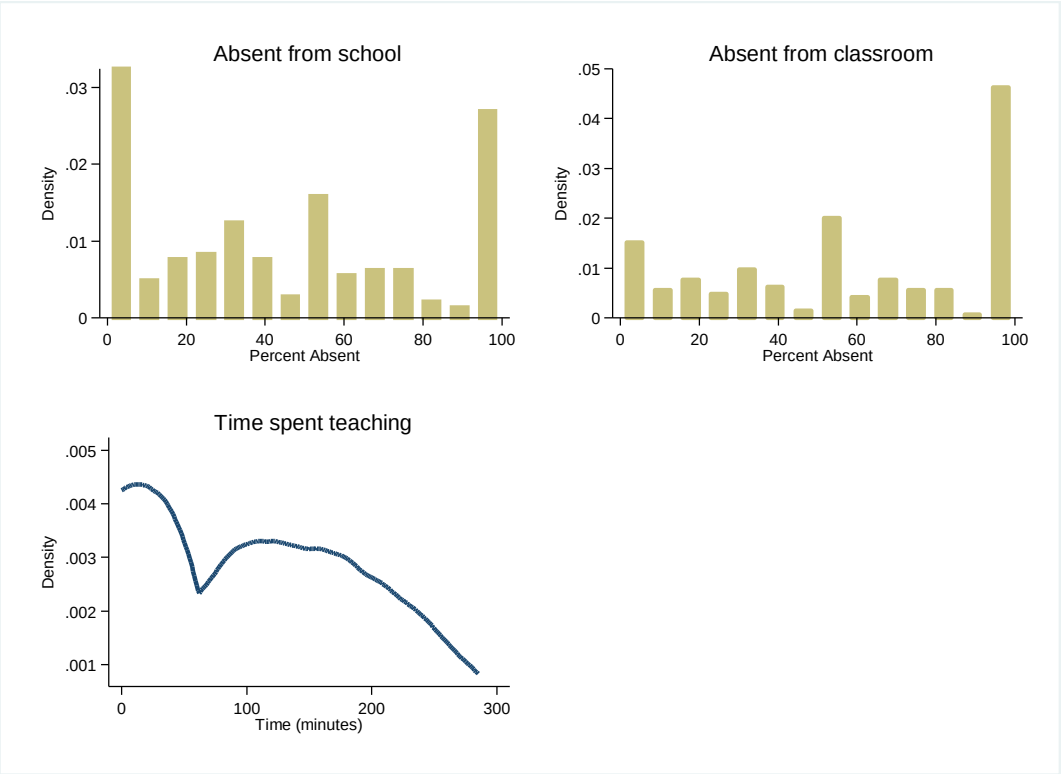
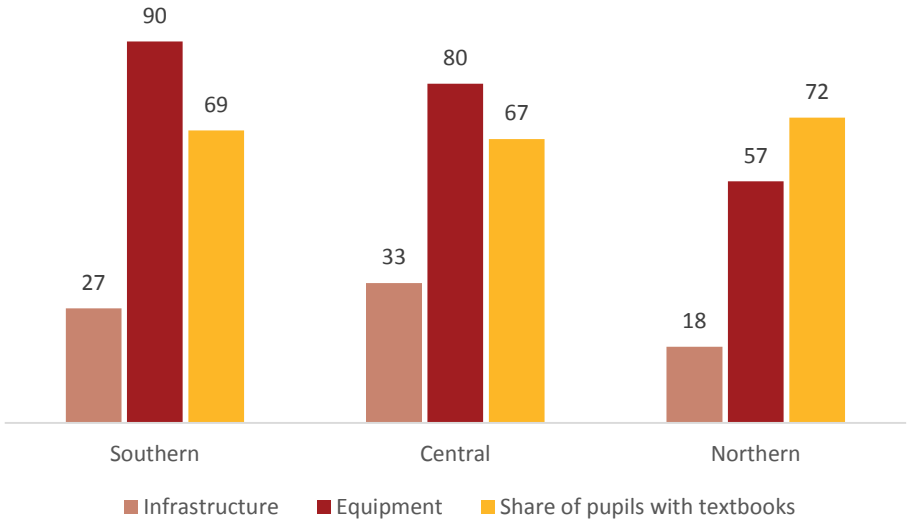


Figure G 7. Infrastructure availability



[25] Definition of "density": In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value. The probability of the random variable falling within a particular range of values is given by the integral of this variable's density over that range—that is, it is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one."(http://en.wikipedia.org/wiki/Probability_density_function).

Figure G 8. Distribution of the input indicators

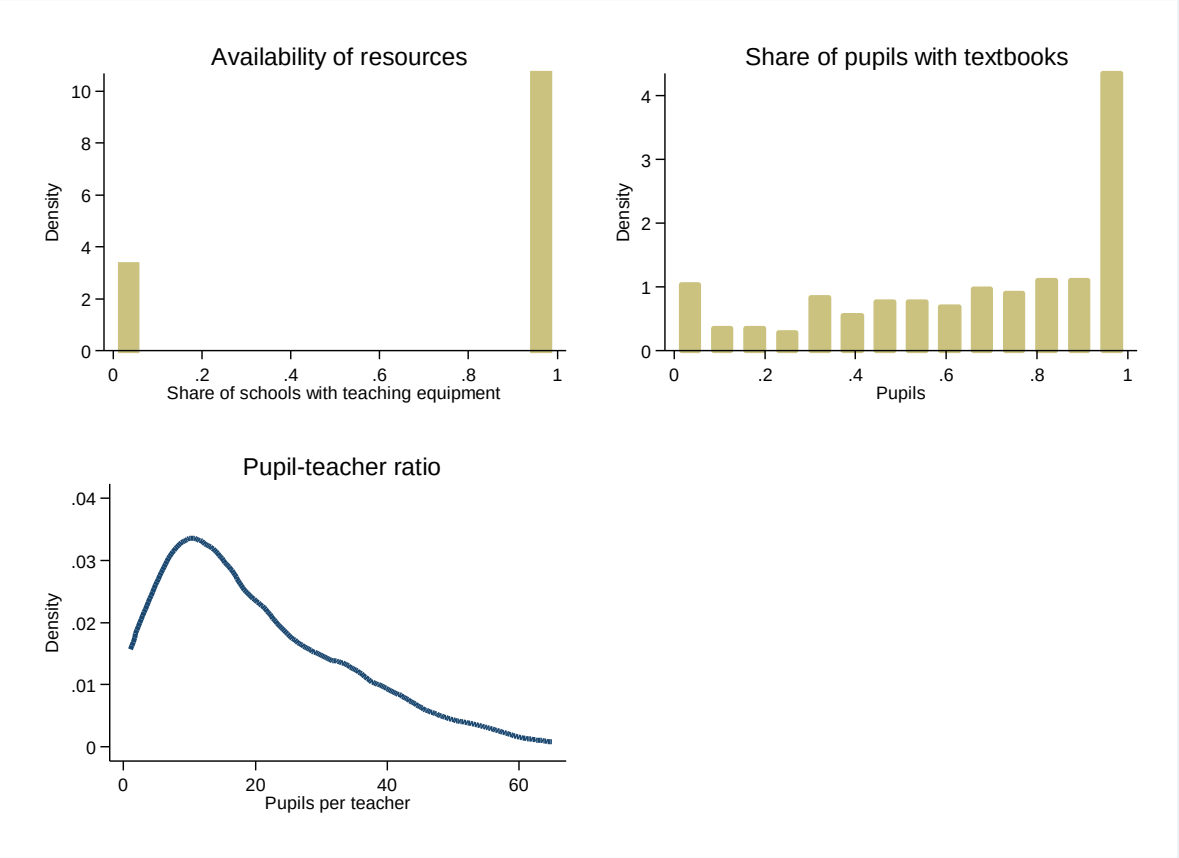


Table G 9. Service Delivery Indicators At-a-glance: Education

	All	Urban Public	Rural Public	Percent Difference (%)	Southern	Central	Northern
Pupil-teacher ratio	21.4	50	17.3	1.3***	33.0	17.3	19.9
Textbook availability (% pupils)	68.1	56.9	70.5	-0.2	66.3	69.6	66.8
Minimum infrastructure availability (% schools)	29.1	34.5	28.0	0.2	33.6	36	13.8
Minimum equipment availability (% classrooms)	76.8	92.3	73.4	0.3***	91.8	81	58.3
Teacher absent from school in 2nd unannounced visit (% teachers)	44.8	33.3	47.9	-0.3***	29.6	48.2	53.0
Teacher absent from class in 2nd unannounced visit (% teachers)	56.2	39.0	60.8	-0.4***	40.9	58.1	67.8
Time spent teaching per day (minutes)	100.9	120.7	96.5	0.2	136.1	97.8	80.0
Minimum teacher knowledge (% teachers)	0.3	0.0	0.4	-1.0	1.6	0.0	0.0

Notes: Weighted means using sampling weight and the sample design. Results for absence rate based on observations from 1006 sampled teachers from 200 schools. Results for time on task based on observations from 200 sampled teachers. Results for teacher knowledge based on observations from 673 sampled teachers. Levels of significance: *** p < 0.01; ** p < 0.05; * p < 0.1.

Table G 10. SDI-At-a-glance (All schools)

	Mozambique+ 2014	Average SDI	Kenya 2012	Nigeria* 2013	Senegal+ 2011	Tanzania 2014	Tanzania+ 2011	Togo 2013	Uganda 2013
Teacher Ability									
Minimum knowledge (At least 80% in language and mathematics)	0.3	**14.6**	40.4	3.7	Not Comparable	15.6	Not Comparable	1.6	11.7
Test score (language, mathematics, and pedagogy)	26.9	**43.0**	57.1	32.9	Not Comparable	46.5	Not Comparable	35.6	42.7
Teacher Effort									
School absence rate	44.8	**18.6**	14.1	13.7	18.0	15.0	23.0	20.5	26.0
Classroom absence rate	56.2	**39.8**	42.1	19.1	29.0	46.6	53.0	35.8	52.8
Scheduled teaching time	*4h 17min*	***5h 34min***	*5h 37min*	*4h 53min*	*4h 36min*	*5h 54min*	*5h 12min*	*5h 29min*	*7h 18min*
Time spent teaching per day	1h 41 min	**3h 02min**	2h 49min	3h 26min	3h 15min	2h 59min	2h 04min	3h 29min	3h 18min
Availability of Inputs									
Observed pupil-teacher ratio	21.4	**40.4**	35.2	21.6	34.0	40.5	74.0	29.7	47.9
Share of pupils with textbooks	68.1	**37.1**	48.0	38.2	Not Comparable	25.9	Not Comparable	68.5	5.0
Minimum equipment availability (90% with pencils and notebooks)	76.8	**60.5**	78.8	54.8	Not Comparable	62.0	Not Comparable	26.4	80.6
Minimum infrastructure availability	29.1	**38.1**	59.5	18.5	Not Comparable	36.6	Not Comparable	22.3	53.7
Pupil Learning									
Test Score (out of 100) (language, mathematics)	20.8	**49.6**	72.0	32.2	Not Comparable	49.5	Not Comparable	45.7	48.6
Language test score	18.7	**49.5**	75.4	31.4	Not Comparable	48.2	Not Comparable	45.5	47.1
Mathematics test score	25.1	**47.3**	59.0	31.9	Not Comparable	57.6	Not Comparable	44.6	43.4

Notes: a. Values for Nigeria are the weighted average of the four states surveyed, namely Anambra, Bauchi, Ekiti, and Niger.
b. These numbers reflect the updated SDI methodology. More information can be found on www.SDIndicators.org.
c. Full definitions of indicators in Annex B.
d. In Mozambique, Senegal, and Tanzania 2011 (round 1) only public schools were surveyed.

VIII. REFERENCES

Caillaud, F. (2015). International Development Association Project Paper on a Proposed Additional Grant to the Republic of Mozambique for an Education Sector Support Project.

Dobbie, W., & Fryer, Jr., R. G. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, 5(4), 28–60.

Duflo, E., Dupas, P., & Kremer, M. (2011). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101(5): 1739-74.

Filmer, D., Molina, E., & Stacy, B. (2015). Teacher knowledge, behavior and student achievement. Evidence from 5 countries in Africa.

Hanushek, E. A. (2006). "School Resources", In Eric A. Hanushek and Finis Welch (Eds.), *Handbook of the Economics of Education*, Volume 2, pp. 865-908.

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

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

Johnson, D., Cunningham, A., & Dowling, R. (2012). "Draft final report, Teaching Standards and Curriculum Review", background document for the SDI.

Lavy, V. (2010). Do Differences in Schools Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *National Bureau of Economic Research, Inc.*

Lavy, V. (2011). What Makes an Effective Teacher? Quasi-Experimental Evidence. *National Bureau of Economic Research, Inc.*

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on pupil achievement: Evidence from within-teacher within-pupil variation. *Journal of Development Economics*, 99(2), 486–496.

Ministry of Education and Human Development, Republic of Mozambique. (2012). "Education Strategic Plan 2012-2016". Maputo, Mozambique: Author.

"Probability density function." Wikipedia. Wikimedia Foundation, Inc., 14 Feb. 2015. Web. 3 Mar. 2015 <http://en.wikipedia.org/wiki/Probability_density_function>

World Bank. (2004). *World Development Report 2004: Making Services Work for Poor People*. World Bank. Washington, DC.