Three Essays in the Economics of Education

Oswald Koussihouede

Université Gaston Berger

Ecole Doctorale des Sciences de l'Homme et de la Société
Unité de Formation et de Recherche
en Sciences Economiques et de Gestion

Three Essays in the Economics of Education

Thèse présentée et soutenue publiquement le 13 Mars 2015
Pour l'obtention du grade de Docteur en sciences économiques

Par:
Oswald Koussihouèdé

Jury:
Mouhamadou Fall, Président, Université Gaston Berger, Sénégal
Tanguy Bernard, Rapporteur, Université de Bordeaux, France
Christian Monseur, Rapporteur, Université de Liège, Belgique
Costas Meghir, Encadreur, Université Yale, Etats-Unis
Adama Diaw, co-Encadreur, Université Gaston Berger, Sénégal

L'Université Gaston Berger n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur(s) auteur(s).

To Rachel-Sylviane, Emmanuelle and Théana,

Acknowledgments

I would like to acknowledge everyone who made this project possible. I could not have accomplished this without the support of my advisor, Professor Costas Meghir, and of my co-advisor, Professor Adama Diaw. Thanks for helping guide my path to the club. My recognition also goes to Tanguy Bernard, Christian Monseur and Mouhamadou Sall who kindly accept to be part of the thesis committee.

I have benefited from the valuable comments of participants at the research seminars organized in Abomey-Calavi (Benin) by the Institute for Empirical Research in Political Economy and in Kinshasa (Democratic Republic of Congo) by the Higher Institute of Statistics. Suggestions at the training workshops organized by the African Growth Development Policy Modeling Consortium in Dakar (Senegal) were very helpful.

I thank the coauthors of the first chapter for both beneficial collaboration and useful discussions.

I am very appreciative to Bruno Abilou, Printys Assèdé and his family, Norbert Assogba, Mame Cheikh Diop, Susan Godlonton, Eudes Hounkpodoté, Hilaire Hounkpodoté, Jacques Boureima Ki (General Secretary of CONFEMEN), Guy Lacroix, Nathalie Lahire, Alessandra Mattei, Samba Mbaye, Damien Mèdédji, Meskerem Mulatu (Lily), Alain-Patrick Nkengne Nkengne and Bassile Zavier Tankeu for their encouragements, thoughtful comments, suggestions or other help provided at various stages of this project. I remain responsible for all errors and views expressed in this document.

Word cannot express the gratitude I owe my whole family for always showing me kindness and support during these years.

To my father, the one who taught me that life is hard, so I have to work harder, I owe my deepest esteem.

To Rachel-Sylviane, Emmanuelle and Théana, who, during these long years, kept asking me when I was going to be finished – this round is completed!

Contents

Dedications **iii**

Acknowledgments **iv**

List of Tables **vii**

List of Figures **xi**

General Introduction **1**

1 Decentralizing Education Resources: School Grants in Senegal **9**

1.1 Introduction 9

1.2 Description of the programme and evaluation 11

1.3 Data and balance 14

1.4 Empirical approach and inference 20

1.5 Results 21

1.6 Conclusion 39

2 Estimating the Causal Effect of School Size on Educational Attainment **40**

2.1 Introduction 40

2.2 School size in the learning process 43

2.3 Data and attrition patterns 44

2.4 School size and school quality in Senegal 45

2.5 Empirical approach 47

2.6 Results and discussions 57

2.7 Sensitivity analyses 68

2.8 Conclusion 76

3 An Extended Specification of the Three-Level Linear Model **79**

3.1 Introduction 79

3.2 A review of multilevel modelling of test scores on PASEC's data 81

3.3 Data and construction of variables 85

3.4 Empirical approach 87

3.5 Results and discussions 92

3.6 Conclusion 104

General Conclusion **105**

Bibliography **108**

Appendix A Supplement to Chapter 1 **127**

Appendix B Supplement to Chapter 2 **158**

Appendix C Supplement to Chapter 3 **208**

List of Tables

1.1 *Baseline Descriptive Statistics and Balance, Grade 2* 17

1.2 *Programme Impacts on Grade 3 Test Scores* 22

1.3 *Programme Impacts on Grade 3 Test Scores by Baseline Ability* 30

1.4 *Programme Impacts on Grade 3 Test Scores by Region* 31

1.5 *South-North Differences in Test Performance, Second-Third Grades* . . . 33

1.6 *South-North Differences in Grade 3 Household and Project Characteristics* 35

1.7 *Programme Impacts on School Characteristics by Region, First Follow-up* 37

2.1 *Estimated ATE and ATET with the Doubly Robust Method by Grade* . . 63

3.1 *Sample Size per Country* 86

3.2 *Analysis of Variance on French and Mathematics Learning Outcomes* . . 93

A.1 *Summary Statistics (Mean and Standard Deviation) of a Nationally Representative Sample of Second and Fifth Graders* 129

A.2 *Student Test Scores Sample Sizes and Attrition* 130

A.3 *Baseline Descriptive Statistics and Balance, Grade 4* 131

A.4 *First Follow-up Descriptive Statistics and Balance, Grade 3* 133

A.5 *Differences in Baseline Characteristics among First Follow-up Non-attriters* 134

A.6 *Differences in Baseline Characteristics among Second Follow-up Non-attriters* 136

A.7 *Characteristics of Schools by Baseline Missing Status, First Follow-up* . . 138

A.8 *Programme Impacts on Grade 3 Test Scores, Standardized Coefficients* . . 139

A.9 *Programme Impacts on Grade 3 Test Scores, No Conditioning Variables* . 140

A.10 *Programme Impacts on Grade 5 Test Scores* 141

A.11 *Programme Impacts on Test Scores at Different Grades* 142

A.12 *Programme Impacts on Grade 3 Test Scores by Region, Standardized Coefficients* . . . 147
A.13 *Programme Impacts on Grade 5 Test Scores by Baseline Ability* . . . 148
A.14 *Programme Impacts on Grade 5 Test Scores by Region* . . . 149
A.15 *Programme Impacts on Grade 3 Test Scores by School Size* . . . 150
A.16 *Programme Impacts on Grade 5 Test Scores by School Size* . . . 151
A.17 *Programme Impacts on Grade 3 Test Scores by Wealth Status* . . . 152
A.18 *Programme Impacts on Grade 5 Test Scores by Wealth Status* . . . 153
A.19 *Programme Impacts on Grade 3 Test Scores by Urban and Rural Location of Schools* . . . 154
A.20 *Programme Impacts on Grade 5 Test Scores by Urban and Rural Location of Schools* . . . 155
A.21 *Programme Impacts on Teacher and Household Outcomes, First Follow-up* 156
A.22 *Programme Impacts on Grade 3 Household Characteristics by Region, First Follow-up* . . . 157

B.1 *Mean and Standard Deviation of Locality- and School-level Covariates by Urban and Rural Location of Schools* . . . 159
B.2 *Mean and Standard Deviation of Other Covariates, Grade 2* . . . 160
B.3 *Mean and Standard Deviation of Other Covariates, Grade 4* . . . 161
B.4 *Mean and Standard Deviation of Characteristics of Small and Large Schools* 162
B.5 *Attrition Models for Midline and Endline Surveys by Grade* . . . 163
B.6 *Estimated Impacts of Large School Size under Different Computation Strategies, Grade 2* . . . 167
B.7 *Estimated Impacts of Large School Size under Different Computation Strategies, Grade 4* . . . 168
B.8 *Models of the Probability of Attending a School of Large Size (Propensity Score) using the Full Set of Covariates* . . . 169
B.9 *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Full Set of Covariates, Grade 2* . . . 171
B.10 *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Full Set of Covariates, Grade 4* . . . 173

B.11 *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Full Set of Covariates, Grade 2* 175
B.12 *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Full Set of Covariates, Grade 4* 177
B.13 *Independent Balance Tests: T-statistics from Tests of Equality of Unweighted Means and Weighted Means of Covariates using a Propensity Score Predicted with the Full Set of Covariates* 179
B.14 *Independent Balance Tests: Normalized Differences between Small and Large Schools* 181
B.15 *Models of the Probability of Attending a School of Large Size (Propensity Score) using the Reduced Set of Covariates* 182
B.16 *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 2* 183
B.17 *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 4* 185
B.18 *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 2* 187
B.19 *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 4* 189
B.20 *Independent Balance Tests: T-statistics from Tests of Equality of Unweighted Means and Weighted Means of Covariates Using a Propensity Score Predicted with the Reduced Set of Covariates* 191
B.21 *Effect of Trimming on Sample Size for the Second Survey Written Test Scores* 193
B.22 *Estimated ATE and ATET with the Method of Inverse Propensity Score Weighting by Grade* 194
B.23 *Estimated ATE and ATET for Subpopulations, Grade 2* 195
B.24 *Estimated ATE and ATET for Subpopulations, Grade 4* 196
B.25 *Estimated ATE and ATET by Social Class Level and by Grade* 197
B.26 *Unconditional Quantile Treatment Effects, Grade 2* 198
B.27 *Unconditional Quantile Treatment Effects, Grade 4* 199
B.28 *Unconditional Quantile Treatment Effects on the Treated, Grade 2* 200
B.29 *Unconditional Quantile Treatment Effects on the Treated, Grade 4* 201

C.1 *Descriptive Statistics by Country* 209
C.2 *Empirical χ^2 values derived from likelihood ratio tests comparing 3-level to 2-level models* 212
C.3 *Reduction in Variance by Type of Conditional Analysis of Variance* . . . 213
C.4 *Estimated Models for Cameroon* 214
C.5 *Estimated Models for Chad* 216
C.6 *Estimated Models for Congo* 217
C.7 *Estimated Models for Ivory Coast* 219
C.8 *Estimated Models for Senegal* 221
C.9 *Estimated Models for Togo* 223
C.10 *Determinants of Academic Growth between Grades 2 and 5, French Test Score* 225
C.11 *Determinants of Academic Growth between Grades 2 and 5, Mathematics Test Score* 226

List of Figures

1.1 *Evaluation Design* 13

1.2 *Evaluation Timeline* 14

1.3 *Distributional Impacts on Test Scores at the Beginning of Third Grade, First Follow-up* 25

1.4 *Distributional Impacts on Test Scores at the End of Third Grade, Second Follow-up* 26

2.1 *School Size Density for Grade 2 (left side) and Grade 4 (right side) at School Years 2009/2010 and 2010/2011* 50

2.2 *Propensity Score Histograms* 60

2.3 *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes* 69

2.4 *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes* 72

2.5 *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes* 73

2.6 *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes* 74

2.7 *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes* 75

2.8 *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes* 76

A.1 *Location of Schools in the Project* 128
A.2 *Distribution of Second-Third Grades French Scores by Gender* 143
A.3 *Distribution of Second-Third Grades Mathematics Scores by Gender* . . . 143
A.4 *Distribution of Second-Third Grades Oral Scores by Gender* 143
A.5 *Distribution of Fourth-Fifth Grades French Scores by Gender* 144
A.6 *Distribution of Fourth-Fifth Grades Mathematics Scores by Gender* . . . 144
A.7 *Distribution of Fourth-Fifth Grades Oral Scores by Gender* 144
A.8 *Distributional Impacts on Test Scores at the Beginning of Fifth Grade, First Follow-up* 145
A.9 *Distributional Impacts on Test Scores at the End of Fifth Grade, Second Follow-up* 146

B.1 *Boxplots for the school size variable* 166
B.2 *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test* 202
B.3 *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test* 202
B.4 *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test* 202
B.5 *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables to the Set of Regressors, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test* 203
B.6 *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables to the Set of Regressors, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test* 203

B.7 *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables to the Set of Regressors, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test* 203
B.8 *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test* 204
B.9 *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test* 204
B.10 *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test* 204
B.11 *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test* 205
B.12 *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test* 205
B.13 *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test* 205
B.14 *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test* 206
B.15 *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test* 206
B.16 *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test* 206
B.17 *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test* 207

B.18 *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test* 207
B.19 *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test* 207

C.1 *Correlations of Grades 2 and 5 Test Scores in French and Mathematics* . 211

General Introduction

The consensus on the importance of schooling finds its roots in the macroeconomic and social benefits related to education. Numerous studies have established a link between quality of education, rapid economic growth, and reduction of levels of poverty and inequality (Vandenbussche et al., 2006; Birdsall and Londono, 1997; Barro and Sala-i-Martin, 1995; Mankiw et al., 1992; Lau et al., 1991). Indeed, education promotes high private and social returns (Moretti, 2004) and is correlated with better individual earnings (Mincer, 1958, 1975), while also improving health and choice in reproduction (Psacharopoulos, 1985, 1994; Schultz 1997, 2002; Strauss and Thomas, 1995). Sen (1992, 1999) has widened the definition of poverty, viewing the concept as an incapability of functioning effectively in society. In this respect, the lack of basic skills can be considered a form of poverty. These scientific results, amongst many others, have been instrumental for convincing policy makers that universalizing a quality education is a crucial component of a country's economic and social development. They further justify why researchers do not limit their investigations to the quality of education, but instead also consider educational equity.

In response to the widely-recognized benefits of education, primary school enrolment has rapidly increased in the developing world. Even in the poorest areas of Sub-Saharan Africa (SSA), gross enrolment rates in primary school are now close to 80 percent (see, for example, Glewwe and Kremer, 2006), if not greater. The increase in school enrolment seems to be the *only* real success of African states in the education sector. While enrolment rates are high, many children remain unenrolled, and many who are enrolled still do not complete primary school (Chimombo, 2005). Also, the *Programme for the Analysis of Education Systems of CONFEMEN* (PASEC) reports that girls tend to learn systematically less than boys in most countries assessed, reflecting an endemic lack of equity in many education systems in SSA countries. Other examples relate to rural areas that are almost always less endowed than urban zones and consequently exhibit lower performance (see PASEC, 2010). In countries evaluated by the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), children from the richest 25 percent of families are ten times more likely to get better results on reading tests than those from the poorest 25 percent of households. There is also widespread evidence that the quality of education in developing countries remains very low. Consequently, it is not surprising that students in these areas exhibit low achievement rates. The observed low achievement relates to conditions such as school management or school effectiveness (Fuller, 1987).

The importance of school management, education financing and school effectiveness in the promotion of students' learning outcomes is widely acknowledged and well-known to policy makers. While policy makers increase education funding,[1] it is not clear how school management or school effectiveness are dealt with. Furthermore, apart from these operational issues, methodological concerns have plagued educational research. A first methodological issue relates to the computation of standard errors in complex sampling designs, which leads to invalid inferences. The confusion of classroom effects with school effects is a second important methodological issue. An example of research that suffers from the problem of standard errors computation is the work of Diop (2011) when it comes to studying the effect of teachers' characteristics on students' achievement. This study, in addition to other disputable methodological choices, fails to account for the actual ways in which students were assigned to classrooms and thus was conducted as if the whole sample was completely random. Other studies plagued by the same drawback of invalid statistical inferences are Coleman et al. (1966), Jencks et al. (1972) and Rutter et al. (1979). Goldstein (1997) offers a discussion of these cases.

The second methodological issue we raised here and that similarly represents a limit of previous studies is the confusion of classroom effects with school effects. This problem arose mainly because of sampling limitations. For a given level, when only one classroom is surveyed or is available in a single school, it is impossible to explicitly account for the classroom effect in a statistical model. Researchers have therefore relied on two-level models instead of three-level models when linking test scores to any relevant explanatory

[1] From an average of 2.87 percent in 2000, governments expenditures on education represented 4.37 percent of the Gross Domestic Product (GDP) in 2010. These calculations are based on data collected at *http://www.uis.unesco.org/Education/Pages/education-finance.aspx* and include the following countries: Benin, Burundi, Cameroon, Chad, Guinea, Mali, Niger and Senegal.

factor. This has been the case for more than two decades of learning assessments run by PASEC. Conclusions reached on the basis of models that ignore the classroom level are likely to be biased, if not inexact. In this respect, Hill and Rowe (1996) find that, when classrooms are included as a level between the student and the school, the classroom-level variance is larger than the school-level variance and the latter is generally reduced to a small value. Therefore, they argue that school effectiveness research should be organized much closer to the classroom than the school.

Our description presents two dimensions of the issues related to education in SSA: first, there are the operational aspects that have not yet come to a satisfactory end, and, second, there is the assessment mechanism of operational actions which suffers from methodological issues. This dissertation tackles both the operational aspects and the methodological concerns mentioned above. More precisely, we propose three empirical essays that are developed in two broad areas of the economics of education. These areas are school-based management for the first essay and school effectiveness for the second and third essays. The general structure of the dissertation matches each essay to a chapter.

The first chapter, titled ***"Decentralizing Education Resources: School Grants in Senegal"*** and co-authored with Pedro Carneiro,[2] Nathalie Lahire,[3] Corina Mommaerts[4] and Costas Meghir,[5] is built upon a randomized evaluation that provides empirical evidence on whether a school grant programme, one which has been running for several years, is able to raise learning outcomes. The second chapter is a quasi-randomized study and is titled ***"Estimating the Causal Effect of School Size on Educational Attainment"***. These two papers use three rounds of data collected in Senegal between 2009 and 2011 with World Bank funding for the purpose of the aforementioned randomized evaluation. The third paper of this dissertation uses learning assessments data provided by PASEC and is titled ***"An Extended Specification of the Three-Level Linear Model"***. This chapter does not cover Senegal only; it also addresses five other countries located in western (Ivory Coast and Togo) and central (Cameroon, Chad and Congo) Africa. The second and third chapters of this dissertation are individually-authored. We describe in the following lines the context of the research and some findings for each chapter.

[2]University College London, London, United Kingdom.
[3]The World Bank, Washington, United States of America.
[4]Yale University, Connecticut, United States of America.
[5]Yale University, Connecticut, United States of America.

As we just set, our first research paper looks at the impact of a site-based initiative in the context of Senegal. The idea of site-based or school-based management (as it is popularly called in the educational research world) goes back many decades. School-based management is the decentralization of authority from the central government to the school level (Caldwell, 2005). According to Malen et al. (1990), “School-based management can be viewed conceptually as a formal alteration of governance structures, as a form of decentralization that identifies the individual school as the primary unit of improvement and relies on the redistribution of decision-making authority as the primary means through which improvement might be stimulated and sustained” (p. 290). The literature on this topic discusses trade-offs without a clear consensus on the superiority of centralization or decentralization in the provision of public services. While the primary arguments in favor of decentralization are that it carries decision-making closer to beneficiaries (and consequently reduces information asymmetries) and improves the accountability of local authorities, decentralized systems can deteriorate the delivery of public goods when local decision-makers are less technically able than national governments (Smith, 1985) or if resources end up being captured by local authorities (Bardhan and Mookherjee, 2005).

Senegal initiated its decentralization process in 1972. The process was reinforced in 1996 with the elevation of the regions to the legal status of local authorities with the same financial autonomy as municipalities and rural communities (Senegal, 1996). Starting in 2000, the government emphasized school-based management as a way to push decentralization forward. We examine the impact of school grants that have been used in Senegal for several years and by several donors as a tool for getting resources closer to the final destination. These grants are based on the premise that school-level actors would be better able to identify the particular deficiencies in the school and the most workable solutions. The overall goal of these grants is to contribute to improvements in the quality of education, often by focusing on the overall physical environment, but sometimes by attempting to address pedagogic issues. Under the first phase of the Senegalese government’s ten-year education and training development programme (French acronym PDEF (Programme Décennal de l’Education et de la Formation)), approximately 1,500 school grants were financed overall, with about 1,000 financed under the International Development Association (IDA) credit during the 2004-2005 school year at a cost of 1.5 million CFA Francs per school (then about US$ 3,190). These grants were implemented well from a financial standpoint, and the internal evaluation of their outcomes by the government appears positive. Despite the general acceptance by the Senegalese government and donors alike that school projects have an important role to play, and despite the considerable resources allocated in recent years to this initiative, a rigorous study has never been conducted to provide the necessary evidence on the effectiveness of these grants. Therefore, the evaluation that follows provides the much-needed evidence for the cohort of projects financed under IDA operation from 2009 through 2011. More broadly, the evaluation provides evidence on whether funding schools through project initiatives designed to match their self-stated needs improves student performance in both early and later grades of primary school.

Prior evidence (e.g. Hanushek et al., 2013) suggests that increasing school autonomy, especially in the decisions regarding academic content, has a negative impact on school achievement in low-income countries. However, in countries with higher incomes, greater autonomy on academic content, staff and budget has a positive impact on student achievement. Likewise, Galiani et al. (2008) find that decentralizing education services from the federal government to the provincial governments has positive effects in secondary schools in Argentina, but only among well-managed localities. Conversely, an analysis based on longitudinal data collected in Chile reveals that national standardized test scores declined during decentralization (Prawda, 1993). A review by Leithwood and Menzies (1998) examined 83 empirical studies of school-based management and reported both positive and negative effects on students and teachers. We find large, positive, and statistically significant effects on test scores one year after the start of the intervention, especially for girls who were first exposed to school grants when they were in second grade. We also detect positive effects of the programme for boys from rich families. The effects are larger for schools in the South of the country, where the projects tended to focus on training human resources (teaching and management), than in the North, where they focused more on the acquisition of school material (e.g. textbooks/manuals). We do not observe similar programme impacts for children in other grades.

The second chapter investigates the effect of school size on learning outcomes by adopting methodologies of quasi-experimentations in the school effectiveness paradigm. Sun et al. (2007) suggest that the objectives of research on school effectiveness are two-fold: the first goal is the identification of factors associated with effective schools, and the second is the documentation of differences between education outcomes in these schools. According to the same authors, there is still a dispute over the educational outcomes to measure in many areas of education research. Nonetheless, test scores or examination results obtained during formal assessment are of central concern. The conceptualization we consider here is proposed by Beare et al. (1989), who define an effective school as one that accomplishes its specific objectives, especially improving learning outcomes. We note obvious discrepancies in the characterization of the concept amongst researchers. Indeed, Cheng (1996) analyzed it from the rationale of production that relates output to input, Willms (1992) related it to growth in students' achievements, and McGaw et al. (1992) suggest the consideration of more than academic achievement.

While a lot is known about the causal impact of class size on students' achievement (e.g. Angrist and Lavy, 1999; Konstantopoulos, 2008; Shin and Raudenbush, 2011), minimal research has focused on the relation of school size to students' achievement. The study of the causal effects of school size on academic achievements falls under the umbrella of school effectiveness since school size represents a dimension of school quality, creating a learning environment to which students must react. Our investigation is dictated by the fact that not only is the broad literature inconclusive as to whether smaller or larger schools are preferable for education systems, but also most of the research conducted in relation to the effects of school size on learning outcomes are correlational analyses. These studies have mostly been carried out in developed countries, which face education issues unlike the challenges being tackled in low-income countries. This state of facts renders the results from existing studies on school size not necessarily applicable to growing economies in general, and especially those located in SSA. Whether school size causally affects learning outcomes in this area of interest was until now an open question. To the best of our knowledge, this research is the first of its kind to look at the potential effects of school size in the context of a developing economy in SSA.

This investigation is important for the policy agenda. As in many other countries, Senegal made progress expanding access to education over the last decade, but quality has deteriorated (DeStefano et al., 2009) as enrolment increased. Not only has quality not improved as enrolment has increased, but the Senegalese government's response to the rising demand for education has been the provision of physical facilities. A question of interest is whether policy makers should construct more schools and keep them small or instead allow schools to be as large as affordable given the demand for education. This paper provides a tentative answer to the question.

The existing literature examining the impact of school size in developing countries typically relies only on correlation analyses (e.g. Crispin, 2011; Walberg and Walberg, 1994), and very few studies have sought to examine the causal relationship between school size and learning outcomes. Kuziemko (2006), Bloom et al. (2010), and Schwartz et al. (2011) suggest negative effects of large schools, whereas Wyse et al. (2008) found no effect at all. Assuming conditional randomization, we mainly use a doubly robust linear regression and quantile regressions as identification strategies. We also conduct extensive robustness checks that suggest long-term and negative effects of large school size on fourth graders and share our intuitions as to how large school size might matter. Furthermore, we estimate 470-500 students to be an "*ideal range*" for school size for the Senegalese context based on the effect of large school size on the overall performance of students.

As noted above, the third and final paper addresses a methodological issue in the research on school effectiveness. We propose a representation of a statistical model that better accounts for complex sampling designs and is more likely to produce valid inferences. In this way, we attempt to disentangle school-level effects from classroom-level effects in a multilevel model. To achieve this, we extend a methodology proposed by O'Dwyer (2002) to learning assessments data that have initially been analyzed either with a classical linear regression model that accounts for clustering at the school level or with a two-level model in the best case. This extension is made possible by means of a within-school grade equivalency procedure which allows us to compare students in grade 2 with those in grade 5 on a single metric for both grades. Students are thereby compared on this single common metric, irrespective of their initial grade. The previous models do not account for the within-schools between-classrooms variation of learning outcomes because the analyses were conducted by grade, and only one grade 2 classroom and one grade 5 classroom are surveyed in participating schools. Yet, the consequences of the omission of a level in the modelling are widely discussed in the literature (e.g. Van Landeghem et al., 2005; Van Den Noortgate et al., 2005; Moerbeek, 2004). The grade equivalency procedure, introduced in the literature by O'Dwyer (2002), is shown to possess the potential to be a tenable solution for separating the within-schools between-classrooms variance from the between-schools variance. A simple decomposition of the variance of learning outcomes over the three dimensions of students, classrooms, and schools shows that the classroom-level variance is sizeable and even significant. Competitive analyses of variance where only two levels are nested in the models fail to perform better than the decomposition of variance where we consider three dimensions. We apply the proposed specification to study students' academic growth over both a period of one year and a longer period of three years. The results of the first application are counterintuitive in some cases. For instance, in Chad and Senegal, we find that rural students outperform urban students. In Congo, we find that students taught by community teachers are the best, everything else being equal; they exhibit better outcomes than students taught by private school teachers, contractual teachers and teachers who are government employees. From the second application, we learn that students' academic progress is more rapid in rural settings and in schools led by female principals. This academic progress rate is also positively correlated with the school's average socioeconomic status, as well as a change in this variable. Conversely, we find that both the principal's experience level and the existence of a management committee are associated with slower academic growth between grades 2 and 5.

Chapter 1

Decentralizing Education Resources: School Grants in Senegal[1]

1.1 Introduction

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

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

We study a school grant programme in Senegal, which was developed to decentralize at least a small part of the country's education budget. Through this programme, every elementary school in Senegal could apply for funds for a specific school project that seeks

[1]This chapter is a joint work with Pedro Carneiro, Nathalie Lahire, Corina Mommaerts and Costas Meghir.

to improve the quality of learning and teaching, with the best proposals being selected through a competitive process. The maximum amount a school could receive for a project amounted to 1,500,000 CFA Francs (approximately USD$ 3,190).

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

Since we examine the impact of the intervention across different tests and different groups of students, for inferential purposes, we implement a recently developed inference procedure (the step-down approach) proposed by Romano and Wolf (2005). This procedure controls the probability of falsely rejecting at least one true null hypothesis, and improves upon more conservative prior methods for multiple hypothesis testing such as the Bonferroni procedure. We show that our main conclusions survive and are unlikely to be due to false rejections.

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

In theory, decentralizing decision making about school resources is a good idea. In practice, this idea is often dismissed. The main problem is that principals in public schools (or other local decision makers) may not be incentivized to use resources optimally to serve students. Recent work on secondary schools in Argentina and primary schools in the Gambia find positive effects of decentralization of school resources (Galiani et al.,2008;

Blimpo et al., 2014). Meanwhile, cross-country comparisons show negative effects of decentralization for developing countries (Hanushek et al., 2013). Our results indicate that decentralized distribution of resources through school grants can have positive effects on student achievement, and we present suggestive evidence that factors such as teacher quality may have enhanced the impacts.

The paper proceeds as follows. In section 1.2, we describe the school grants programme in Senegal and the evaluation design. In section 1.3, we describe our data. In sections 1.4 and 1.5, respectively, we present our empirical approach and the main results where we examine potential mediating factors through which the impact of the programme may have operated. In section 1.6, we conclude.

1.2 Description of the programme and evaluation

Primary schooling in Senegal consists of six years of education and is funded through a mix of government, foreign aid, and parent resources.[2] Almost all classroom instruction is conducted in French, while the language spoken at home is predominantly not French (only 11 percent of the household interviews were conducted in French). According to the Senegalese Ministry of Education, gross enrolment rates in primary school increased dramatically over the ten years prior to our study, from 67 percent in 2000 to 92 percent in 2009. Despite this large increase in enrolment, in 2009, only 60 percent of students completed primary school. In an effort to increase the quality of primary education, the Senegalese Ministry of Education implemented the school grants programme.

1.2.1 School grants in Senegal

For the past several years, Senegal has used school grants (*projets d'école*) as a tool to bring resources closer to their final destination, based on the premise that school-level actors are in the best position to identify a school's unique deficiencies and the most workable solutions to address them. Beginning in 2009, the Senegalese government made an informed decision to use grants as a means to contribute to improvements in education quality with an emphasis on addressing pedagogic issues.[3] The government

[2]Fees collected from parents represent around ten percent of school funding in 2006 (see table 1.5 in the learning assessment report by the *Programme for the Analysis of Education Systems of CONFEMEN* (PASEC), 2007) and are a non-trivial financial burden on families: around one-fifth of students who dropped out in the first year of primary school did so because of limited financial resources of their parents (World Bank, 2013).

[3]Previous grants were mainly geared towards strengthening the physical environment.

also sought technical and financial support from the World Bank to rigorously evaluate the programme.

Generally, the programme works as follows. The Ministry of Education issues a call for proposals, based on the objectives, available grant funding, priority areas, eligible activities (and sometimes eligible regions), and implementation modalities of the donor or donors (for example, the World Bank or the Japanese International Cooperation Agency). Schools that decide to apply for funding complete a grant application for a school project addressing a particular issue faced by the school. The project is approved by a school committee comprised of teachers, parents, and community leaders. New guidelines required all proposed activities to relate directly to pedagogical activities.

Each grant totaled around 1,500,000 CFA Francs (approximately USD$ 3,190). This can represent a seven percent increase in expenditures per student in a typical school, inclusive of teacher salaries.[4] Another important component of the programme was its role in promoting strong community participation in schools. Grants were prepared by a committee of parents, teachers, and local officials.

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

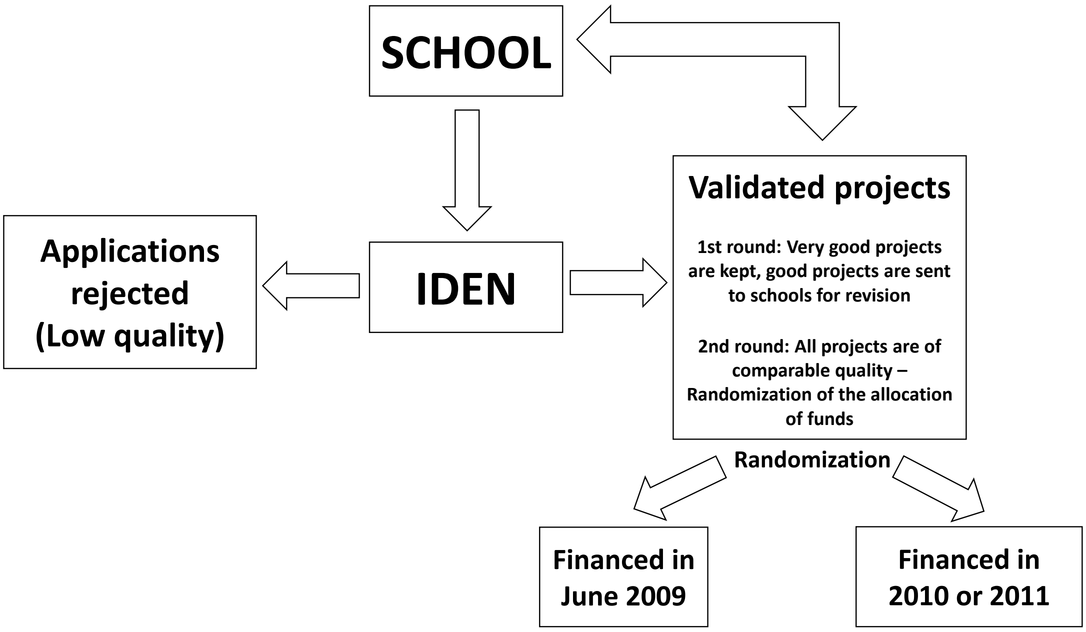
1.2.2 Evaluation design

In the initial stage of this study, all Senegalese schools were eligible to respond to the call for proposals. As illustrated by the diagram below, the IDEN committee ranked the applications, and discarded low quality and ineligible applications. The remaining ones, referred to as "*approved applications*" were grouped into two categories. The first consisted of very good proposals which were considered to be eligible for financing. The second consisted of strong proposals with potential, but which needed revision. These were sent back to schools with comments from the IDEN evaluation committee, then

[4]Based on collected self-reports from principals and teachers in our sample.

re-submitted to the IDEN. Figure 1.1 provides a graphical representation of this process.

Figure 1.1: *Evaluation Design*

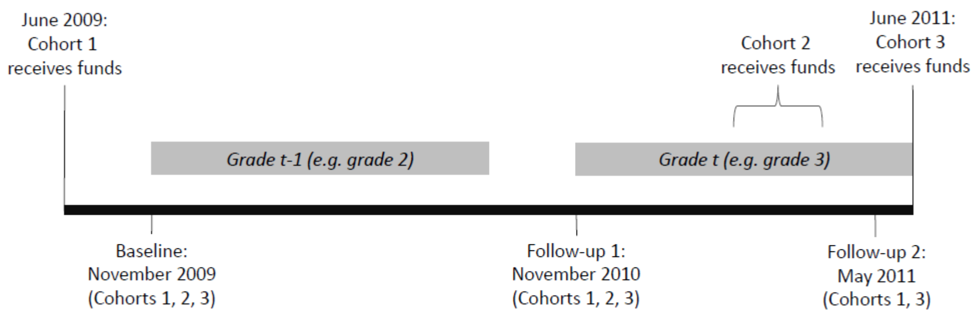


To implement this process, the procedures manual for the projets d'école was amended (relative to versions used for earlier cohorts of school grants) to include the revision of strong proposals needing adjustments. An additional official document issued by the Ministry of Education was circulated throughout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

This process resulted in the selection of 633 projects for funding; the projects' locations are shown in figure A.1. Of these projects, 96 percent included a component to improve French outcomes, 70 percent had a component to improve mathematics outcomes, and 52 percent had a component to improve science outcomes. 82 percent of the projects aimed to build capacity, 63 percent aimed to increase teaching time, and 45 percent aimed to reduce repetition and dropout. The intended beneficiaries of these projects, in addition to students, were the teachers and principal in 84 percent of projects, and the management committee in 29 percent of projects.

For the purposes of the evaluation, these 633 projects were randomly allocated to three funding cohorts. 211 schools were selected randomly to receive funding in the first cohort, at the end of the school year 2008-2009. This funding could only be executed at the beginning of the following school year (October/November). Of the remaining schools, 211 were to receive funding in June 2010, and another 211 were to receive funding in June 2011. In practice, the disbursement of the second round of grants did not occur until the first trimester of 2011. The schools in the first cohort received grants and implemented their projects during this period, while the schools in the second and third cohorts did not receive grants. Yet, schools in the control group accessed financial and pedagogical materials from other sources between the baseline and first follow-up surveys, a fact that prevents us from estimating the full impact of the program. However, the resources they received are pretty low compared to the financial and pedagogical materials afforded by the grants in the treatment group. Thus, we can still use schools in the second and third cohorts as a comparison group for the schools in the first cohort. The school year runs from October/November through June, allowing us to compare the first cohort to both the second and third cohorts for the 2009-2010 school year and the first cohort to the third cohort for the 2010-2011 school year (see figure 1.2).

Figure 1.2: *Evaluation Timeline*



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

1.3 Data and balance

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

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

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

The first follow-up took place at the start of the school year in November 2010. The second follow-up took place at the end of that same school year in May 2011, when we re-surveyed and tested the same children who were tested at baseline and first follow-up. Students were given the same mathematics, French, and oral test across all waves. In addition, in the second follow-up, we administered the Peabody Picture Vocabulary Test (PPVT) to children and their mothers. As with the first follow-up, we administered written tests in mathematics and French to a random sample of six students in each of grades 2, 4, and 6, and oral tests to three of these six children. We did not collect general school and classroom information in the second follow-up.

Of the 633 schools, split *randomly* into three cohorts of 211 schools each, we sampled 525. We were able to contact 478 schools at baseline (among which 447 were successfully surveyed), 528 at first follow-up[5] at first follow-up (among which 517 were successfully surveyed) and 340 at second follow-up (among which 325 were successfully tested and surveyed).[6]

[5]Although we plan to sample 525 schools, three schools (one treatment and two controls) from the remaining 108 schools (planned to be funded but not surveyed) were accidentally contacted by enumerators.

[6]See table A.2 in the appendices for the corresponding number of student-level observations and attrition.

The schools that were not included in the baseline were inaccessible either due to inclement weather or rebel activity in the South. While this may have impacted the representativeness of the baseline sample, it did not affect the balance as accessibility was not correlated with treatment status, as we will report later. Due to budgetary constraints, in the second follow-up, we dropped schools in the second cohort and ended up with a total sample of 352 schools, of which 325 schools were successfully surveyed and tested. Since cohorts were randomly allocated, this did not introduce bias.

The first column of panel A of table 1.1 shows means and standard deviations of baseline test scores for female students in grade 2. The third column shows the same results for boys (analogous results for grade 4 are in appendix table A.3). We can see that the French, mathematics, and oral tests were appropriately targeted to student grade level, as mean scores (calculated as the proportion of correct responses on the exam) were around 30-40 percent. The same tests were administered at first follow-up, so these scores allowed room for noticeable improvement. The fourth row corresponds to an index of the three tests (which is the first principal component of these three tests, standardized to have mean zero and unit variance).

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

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

The second and fourth columns in panels A and C show the differences in test scores and household characteristics between control and treatment at baseline, for both male and female students. The second column of panel B shows differences in the characteristics of treatment and control schools. All, but one difference (parental involvement in school), are insignificant. Thus, treatment and control are very well balanced.

In the tables A.5 and A.6, we show the differences in baseline characteristics between treatment and control among students who did not leave the sample between the baseline and the first follow-up or second follow-up. The sample is similarly balanced as our main sample (see below).

Table 1.1: *Baseline Descriptive Statistics and Balance, Grade 2*

Panel A: Test Scores

	Females		Males	
	Control	Differences	Control	Differences
Percentage of correct answers: French	0.413	0.001	0.435	0.010
	(0.219)	(0.018)	(0.225)	(0.019)
Percentage of correct answers: Mathematics	0.351	0.006	0.382	0.001
	(0.223)	(0.018)	(0.234)	(0.020)
Percentage of correct answers: Oral	0.212	-0.011	0.235	-0.009
	(0.165)	(0.016)	(0.174)	(0.018)
Index Score	-0.085	-0.012	0.081	0.012
	(0.952)	(0.096)	(0.991)	(0.104)

Panel B: School and Teacher Characteristics

	All Students	
	Control	Differences
Distance of the school locality to the nearest city (in kilometres)	18.378	0.067
	(25.007)	(2.176)
Locality population (in 100,000s inhabitants)	1.383	-0.035
	(4.400)	(0.459)
The school locality has a health centre	0.709	-0.026
	(0.454)	(0.043)
The school is located in the South of the country	0.185	0.007
	(0.388)	(0.037)
The school has electricity	0.566	-0.013
	(0.496)	(0.048)
Number of teachers	9.678	-0.439
	(4.966)	(0.511)
Number of pupils (School size)	341.113	-28.473
	(252.387)	(25.604)
The school has a library	0.206	-0.079*
	(0.405)	(0.042)
Number of computers	1.281	0.011
	(4.386)	(0.397)
Proportion of female teachers	0.316	-0.010
	(0.235)	(0.023)
Average age of teachers	33.118	0.132
	(4.242)	(0.389)
Percentage of teachers with a Baccalaureate degree	0.413	0.019
	(0.227)	(0.023)
Average experience of teachers	6.558	-0.083
	(3.693)	(0.350)
Teachers had training in the five years preceding the grant	0.474	-0.096**
	(0.499)	(0.049)
Number of manuals in classroom	59.897	-3.173
	(45.183)	(4.581)

Table 1.1: *Continued*

Panel C: Household Characteristics

	Females		Males	
	Control	Differences	Control	Differences
Absenteeism (in days missed the week preceding the survey)	0.209	-0.057	0.128	-0.070
	(1.085)	(0.102)	(0.580)	(0.081)
Proportion of students that participate in housework	0.010	-0.016	0.010	0.003
	(0.100)	(0.014)	(0.099)	(0.009)
Household size	9.365	-0.223	9.301	0.241
	(4.048)	(0.423)	(4.065)	(0.432)
Number of children in the household	5.211	-0.183	5.350	0.277
	(2.568)	(0.269)	(2.665)	(0.288)
Proportion of Heads of household with any education	0.605	0.059	0.586	-0.037
	(0.490)	(0.050)	(0.493)	(0.051)
Proportion of adult females with any education	0.366	0.038	0.363	0.010
	(0.399)	(0.041)	(0.419)	(0.043)
Proportion of literate heads of household	0.622	0.044	0.616	-0.001
	(0.486)	(0.049)	(0.487)	(0.050)
Proportion of literate adult females	0.353	0.025	0.346	-0.003
	(0.392)	(0.041)	(0.413)	(0.042)
Distance of the student's home to school (in kilometres)	0.697	0.035	0.742	0.108
	(0.821)	(0.077)	(0.987)	(0.086)
The student's parents are involved in the school activities	0.369	-0.141***	0.401	-0.003
	(0.483)	(0.050)	(0.491)	(0.051)
Household food expenditures (in 1,000s CFA Francs)	21.806	-1.385	21.828	-1.273
	(16.099)	(1.524)	(14.830)	(1.533)
Expenditures on uniform (in 1,000s CFA Francs)	2.442	-0.378	2.423	0.131
	(1.395)	(0.461)	(0.966)	(0.385)
Expenditures on tuition (in 1,000s CFA Francs)	1.114	0.017	1.107	-0.003
	(1.091)	(0.113)	(1.279)	(0.117)
Expenditures on supplies (in 1,000s CFA Francs)	3.460	-0.151	4.281	0.906*
	(2.475)	(0.267)	(7.614)	(0.497)
The student has a tutor	0.148	-0.007	0.157	0.026
	(0.355)	(0.038)	(0.365)	(0.036)
The student's home has electricity	0.478	-0.038	0.471	-0.016
	(0.500)	(0.052)	(0.500)	(0.052)
The student's home has a television	0.485	-0.025	0.464	-0.082
	(0.501)	(0.052)	(0.500)	(0.053)
The student's home has a modern toilet	0.528	0.012	0.572	0.026
	(0.500)	(0.053)	(0.496)	(0.051)
Land owned by the household (in hectares)	2.234	-0.262	2.508	-0.618
	(3.104)	(0.451)	(3.789)	(0.813)
The interview is conducted in French language	0.114	0.030	0.121	0.049
	(0.318)	(0.029)	(0.327)	(0.031)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard deviations in parentheses in columns 1 and 3.
Clustered standard errors in parentheses in columns 2 and 4.

It is noteworthy that the precision of the difference in test scores is very high, implying a confidence interval of only ±3% around the point estimate, which bodes very well for our ability to detect even small effects of the programme.[7]

As we explained above, some schools were inaccessible at baseline, and thus were only added to the survey in the first follow-up (although they participated in the randomization, and the treatment schools in this group were funded as planned). The exclusion from baseline was unrelated to treatment status, which explains why, nevertheless, baseline schools are balanced. In the appendix table A.4, we present descriptive statistics for all schools including those added at the first follow-up. As we expect, when we compare the characteristics of treatment and control schools which we did not expect to change as a result of the experiment, there is no significant difference, other than possibly in distance from school. However this is just one significant difference among many; jointly there are no differences and this one is very small in magnitude. Hence, whether we look at schools surveyed at baseline (November 2009) or at the first follow-up (November 2010), there is no evidence of imbalances between treatment and control, with respect to their baseline characteristics.

Another concern is that these 633 schools may be fundamentally different from other primary schools in Senegal. In order to be in our sample, schools had to develop a project good enough to be eligible for funding (perhaps after a revision, as explained above). Therefore, they may not constitute a random set of schools in Senegal. In appendix table A.1, we show characteristics of a nationally representative sample of Senegalese households using data collected in 2006/2007 by PASEC, a survey aimed at assessing educational attainment in primary school, and variables that correspond to those in our data. Schools in our sample have fewer students and are more likely to have electricity than the average school in Senegal, but are similar on other measures, including the presence of a health centre in the locality, the number of teachers and their education, and whether the school has a library. Households in our sample are less likely to have electricity or a television in the home, although they are more likely to have a toilet.

[7]With the exception of the index score, in this paper, we chose not to standardize the mathematics, French, and oral scores. The tests were designed to appropriately measure the types of skills taught in the first years of elementary school in Senegal, and looking at the proportion of right answers in this test is a natural way to assess student knowledge in these subjects, and its progress over time. Furthermore, these scores are specific to Senegal, so standardization would not be useful for international comparisons. Even within sample, we show below that the distribution of scores is highly non-normal, so a one standard deviation in test scores does not have the usual meaning. Nevertheless, in the appendices, we replicate our main results using standardized test scores.

Literacy rates are similar. At least, in terms of these variables, our sample does not look drastically different from the average Senegalese primary school.

1.4 Empirical approach and inference

We use a regression based approach to estimate the impacts of the programme. Specifically, the impacts are the estimated Δ_t^k coefficients[8] from the following regression:

$$Y_{ist}^k = \alpha_t^k + \Delta_t^k T_s + X_{is}\delta_t^k + \varepsilon_{ist}^k \tag{1.1}$$

where Y_{ist}^k is the proportion of correct answers in test k, for student i in school s at follow-up t (1 or 2), T_s is a treatment indicator, X_{is} are conditioning variables unaffected by the grant, and ε_{ist}^k is an error term such that $\mathbb{E}(\varepsilon_{ist}^k) = 0$.

Conditioning variables include household size, number of children, whether the head has any education, distance to school, a wealth index,[9] the interview language, and the baseline scores of all tests. Since household interviews were conducted for only a random subsample of students, two-thirds of our sample has missing household characteristics (at random). In order to keep these observations, we assign zeros to conditioning variables if they are missing and include dummies for observations with missing conditioning variables.[10]

We present standard errors that are robust to heteroskedasticity. Since we are testing multiple hypotheses at once, we compute levels of significance for each coefficient using the step-down approach of Romano and Wolf (2005). In this way, we control for the family-wise error rate. The family-wise error rate is defined as the probability of incorrectly identifying at least one coefficient as significant. Thus, our approach is to control for a family-wise error rate of 5 and 10 percent and mark each coefficient that is significant at each of these rates with ** and * respectively.

We also present bootstrap-based 95% confidence intervals. They may exclude zero, pointing to a coefficient that is significant at a 5% level, while the multiple hypothesis

[8]Appendix table A.8 reports results using standardized test scores.

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

[10]Results without conditioning variables are presented in appendix table A.9 and they are almost identical, but of course less precise.

testing adjusted p-value may be above 5% or even above 10%. This is perfectly consistent because the p-value is adjusted for multiple hypothesis testing, while inference based on the 95% confidence interval only corresponds to a 5% level of significance if no other hypothesis is tested.

Finally, testing too many hypotheses at once may reduce power to detect anything significant. We thus test multiple hypotheses in related groups rather than for all effects reported in the paper.

1.5 Results

1.5.1 Impacts on test scores for boys and girls

In table 1.2, we present estimates of the difference between third grade test scores in treatment and control schools, measured at first follow-up (beginning of third grade) in panel A, and at second follow-up (end of third grade) in panel B.[11]

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

Before proceeding to a more detailed analysis of the impacts, we first note that the effects of the programme for third grade students on the entire set of 42 main outcomes we consider[12] are jointly significant with a p-value of 0.02. This conclusion is based on a χ^2 type test with critical values derived using the bootstrap. Thus, the intervention did have an overall significant effect on third grade outcomes.

We start with the results for children who were in third grade at the first follow-up, one year after the intervention started. Panel A of table 1.2 has four columns, one for

[11] Figures A.2 through A.7 show how the distribution of test scores for female and male students changes over time as a result of the experiment.

[12] The outcomes jointly tested include the 3 tests for boys and girls in both follow-ups plus the PPVT. We consider a national sample (14 hypotheses), and a sample split between the North and South of the country (28 hypotheses).

each of the three tests, and one for the aggregate index of all tests. There are two rows. The first presents the estimates for female students and the second presents estimates for males. Each cell displays estimates of Δ_t^k, the corresponding 95% bootstrap confidence interval, and significance stars based on the Romano and Wolf (2005) procedure.

Table 1.2: *Programme Impacts on Grade 3 Test Scores*

Panel A: Beginning of Third Grade, First Follow-up

	French	Mathematics	Oral	Index[a]
Females	0.037	0.031	0.047*	0.217**
	[0.005, 0.068]	[0.004, 0.057]	[0.015, 0.080]	[0.071, 0.358]
Males	0.022	0.024	0.011	0.041
	[-0.009, 0.055]	[-0.004, 0.050]	[-0.024, 0.044]	[-0.100, 0.188]
Observations	2,720	2,718	1,385	1,350

Panel B: End of Third Grade, Second Follow-up

	French	Mathematics	Oral	Index[a]	PPVT[a]
Females	0.043	0.019	0.054	0.246	0.096
	[0.006, 0.082]	[-0.016, 0.053]	[0.007, 0.103]	[0.057, 0.447]	[-0.155, 0.358]
Males	0.026	0.016	0.024	0.078	0.207
	[-0.008, 0.063]	[-0.019, 0.051]	[-0.018, 0.066]	[-0.099, 0.251]	[-0.016, 0.410]
Observations	1,732	1,721	853	826	566

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

There are large impacts of school grants on third grade test scores of girls. This is true across all tests. Impacts range from an increase of over three (French and mathematics) to almost five (oral) percentage points in the proportion of correct answers in each test. These are large effects in light of the means (and standard deviations) of these test scores: 0.54 (0.24) for both written French and written mathematics, and 0.35 (0.22) for the oral test. If we view all of these tests as noisy measurements of one underlying human capital factor, we may improve precision by using the first principal component computed using factor analysis. When we look at this aggregate index of the three tests, the school grant increases third grade school performance of females by 0.22 of a standard deviation.

All of these effects are individually significant (i.e. the 95% confidence interval does not include zero) for girls, but none are for boys, where the effects are much smaller and the index points to no overall improvement. To evaluate the statistical significance of these effects as a group, as explained above, we first carry out a joint χ^2 test for all effects reported in table 1.2 (excluding the index). Jointly these effects are significant at the 2% level. To pinpoint which effects are responsible for the overall rejection of the null hypothesis, we implement a procedure recently developed by Romano and Wolf (2005). This procedure controls the probability of falsely rejecting at least one true null hypothesis, which becomes more important as the number of hypothesis tests increases. The Romano-Wolf approach improves upon more conservative classical methods such as the Bonferroni correction by applying a "*step-down*" algorithm that takes advantage of the dependence structure of individual tests. Based on this procedure, the oral test for girls is significant at the 10% level but none of the other effects are.[13] However, the effect on the aggregate index for girls, which effectively filters out measurement error, is still significant at the 5% level, once we control for multiple testing.[14] Taken together, these results imply substantial, sizable and significant effects on the performance for girls but did not change the overall performance of boys. This is evident in the point estimate for the index for boys which implies just a 4% of a standard deviation improvement in student performance, not statistically different from zero.

Panel B of table 1.2 documents programme impacts measured using test scores at the end of third grade for the same students (the second follow-up). In addition to the three tests and the index described above, column 5 shows the impact on PPVT scores, standardized to have mean zero and unit variance in the sample.

Impacts of the programme for boys are positive but much smaller than for girls, below 2.8 percentage points for all tests considered and in both follow-ups (about 0.08 of a standard deviation for the index). These are not significant by any criterion.

For females, impact sizes are similar to those reported in Panel A of table 1.2, if not larger, with the exception of mathematics. This indicates that programme impacts persisted two years after the grant was disbursed to schools. While the effects survived until the second follow-up, they are not additive, but rather comparable, and we cannot

[13]Here we are including the French, mathematics, and oral outcomes for boys and girls for the first follow-up, overall and by North and South of the country - results to be discussed below - eighteen hypotheses jointly.

[14]Here the joint test includes the index for the first follow-up for males and females, overall and by North and South of the country - i.e. six hypotheses jointly.

rule out that the programme has failed in putting students on steeper learning trajectories. The effects we report for girls are all individually significant, except for the PPVT and mathematics in the second follow-up. Our interpretation of these results is that the programme seems to be improving the performance of girls who, at the start of the programme, were in second grade (and tested at the beginning and the end of third grade).

It is interesting that a relatively small grant is able to improve children's learning outcomes to this extent, especially for girls. By contrast, in Glewwe and Kremer's (2006) survey of the recent literature on the effectiveness of improvements on school resources on students learning in developing countries, there are several interventions that show no significant impact. In developed countries, there are even fewer examples of successful school resources interventions (Hanushek, 2006). However, we note that, based on the step-down p-values, the effects are not significant in the second follow-up.[15]

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

Whether the programme has different effects across the distribution is an important question relating to targeting. In figure 1.3, we show parameter estimates together with their 95% confidence intervals from a quantile regression of the relevant test scores for grade 3 in the first follow-up, on the treatment indicator and including the usual controls. The standard errors are robust to heteroskedasticity and clustered at the school level.

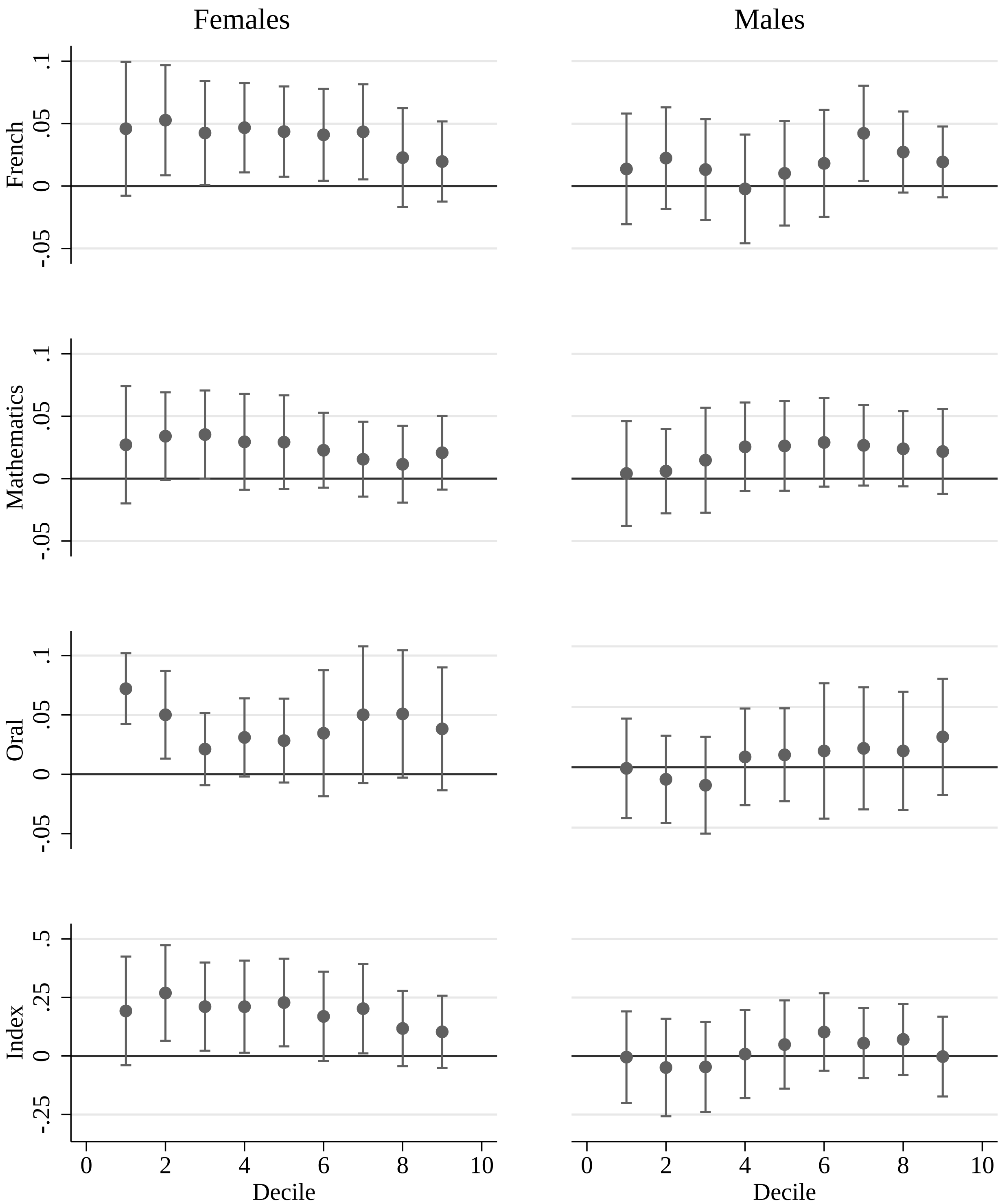
For female students, the effects of the grant are spread over most of the distribution, except at the very top. This is most apparent when we plot the quantile treatment effects for the index. For boys these results confirm no effect along the entire distribution.[16]

A similar picture emerges for the second follow-up in figure 1.4, although the results are unfortunately less precise because of the smaller sample size. Finally, estimates of the grant on PPVT scores occur mainly in the middle of the distribution, for boys. No other effect is observed for boys on the remaining test scores.

[15] Notice that all the individual confidence intervals increase relatively to those in panel A. For this analysis, the sample used is smaller, since we could not survey cohort 2 schools).

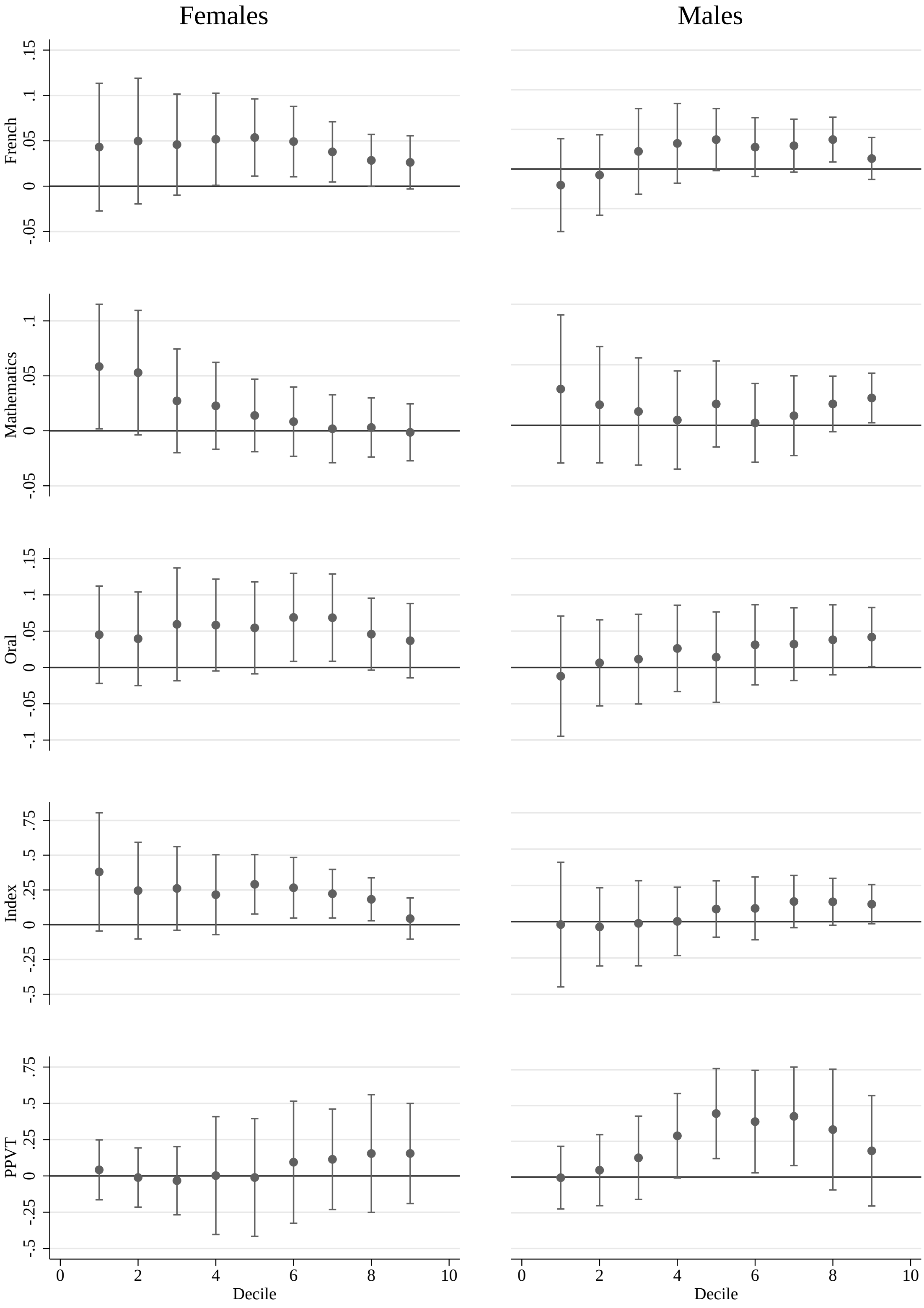
[16] See figures A.8 and A.9 for grade 5 results.

Figure 1.3: *Distributional Impacts on Test Scores at the Beginning of Third Grade, First Follow-up*



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

Figure 1.4: *Distributional Impacts on Test Scores at the End of Third Grade, Second Follow-up*



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

Across survey waves, the two groups of children that do not appear to ever benefit from the grants are boys at the very bottom of the test scores distributions, and girls at the very top. Some of the estimates for boys are even negative.

Finally, we have also estimated the main effects for children who were at grade 5 in the two follow-ups. The impacts, which are numerically close to zero, and statistically insignificant by any criterion, are reported in table A.10.[17] It is remarkable that there are only (individually) statistically significant programme impacts (either positive or negative) on grade 3 tests. The standard errors of the estimates are similar across grades, but the point estimates are quite small across all tests, grades, and survey waves, with the exception of grade 3.[18]

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

Using data from the teachers' questionnaires at follow-up, we investigate whether there were differential impacts of school grants on a few observable investments in 3^{rd} and 5^{th} grades students in panel A of appendix table A.21.[19] Some of the variables we can study are classroom materials (e.g. textbooks/manuals, desks, tables, etc), and teacher training. We find no differential impact of the programme in any of these. When we examine other classroom characteristics or teacher behaviors, the only interesting difference to report concerns student (mis-)behavior in the classroom. While in third grade there was a positive impact of the programme on student behavior as measured by the number of times a day a teacher needs to demand silence, in fifth grade, there was a negative impact of the programme on student behavior measured by this variable, and by the number of times a teacher has to punish a child for impolite behavior.

It is also possible that parents believe that investments in the early grades are more

[17]Appendix table A.11 reports these effects without controls, as well as the impacts of school grants on test scores of children in other grades for which we collected test scores but not other data.

[18]Therefore, the lack of statistically significant results in grade 5 (but not in grade 3) does not appear to be due to a lack of power. If the point estimates for grade 5 were as large as those for grade 3 it is likely that we would be able to reject that they were statistically equal to zero. When designing our study we anticipated that with our sample we would be able to detect programme impacts of between 0.20 and 0.30 standard deviations, which is in line with what we find.

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

productive than investments in later grades. If that were true, it could happen that parents decreased their home investments (if they were substitutes with school investments) more in later grades than in earlier grades in response to an increase in school resources. However, once again, there is no evidence that this took place, at least in terms of observable parental investments (see panel B of appendix table A.21).

1.5.2 Heterogeneous impacts

There are important differences in the impact of school grants on the test scores of boys and girls. Here we consider two additional characteristics by which the impact of the school grants may plausibly differ: prior ability and region.

For baseline ability, we convert corresponding baseline test scores into a "*high*" (above median) or "*low*" (below median) binary variable. However, as we mentioned above, several schools were missing at baseline. In appendix table A.7, we show that missing schools at the baseline are mainly in the South and that they display worse student performance in the first follow-up than comparable non-missing schools. It is noteworthy that they are not disproportionately control or treatment schools.

The second distinction that turns out to be important, and that we emphasize here, is between schools located in the most southern regions in Senegal (Ziguinchor and Kolda), and schools in the rest of the country. While this distribution may at first sound ad hoc, the motivation for considering these regions separetly is because Ziguinchor and Kolda are the much poorer southern regions (Ministry of Economy and Finance of Senegal and World Bank, 2004) and have been beset by problems related to rebel activity.

Since our larger estimates of programme impacts were for students in 3^{rd} grade, who were first exposed to the programme in 2^{nd} grade, we focus this analysis of heterogeneous impacts on them. A similar analysis performed on the test results of students in 5^{th} grade did not produce evidence of any programme impacts for this set of students (see tables A.13 and A.14).

The regressions we run to construct tables 1.3 and 1.4 extend equation 1.1 to include an interaction between the treatment variable T_s and a predetermined variable W_{ist} (baseline ability or region to start with):

$$Y_{ist}^{k} = \alpha_t^k + \Delta_t^k T_s + \delta_t^k \left(T_s * W_{ist}\right) + \psi_t^k W_{ist} + \varepsilon_{ist}^k \tag{1.2}$$

Tables 1.3 and 1.4 have two panels, corresponding to W_{ist} ="*Baseline ability*" and W_{ist} ="*Region*", respectively. Each panel reports the treatment effect for each W_{ist} (δ_t^k)

and the 95% bootstrapped confidence interval. The final row reports the p-value of the test of equality of the two mean treatment effects.

The first panel of table 1.3 shows that the impact of school grants on 3^{rd} grade French scores is especially large for girls with a high level of baseline ability in French. If investments in skills are complementary over time, they will be more productive for those with high levels of skill to start with. There are several education interventions that share this characteristic.

There are also several education interventions that benefit mostly girls (e.g. Burde and Linden, 2013; Garibaldi et al., 2012; Jackson, 2010; Angrist et al., 2009; Angrist and Lavy, 2009; Dynarski, 2008; Anderson, 2008; Kling et al., 2007; Hastings et al., 2006). It is much less common to find programmes that affect boys alone. Again, this may be related to the skills that girls bring to elementary schools, such as discipline, patience, and higher levels of maturity overall, which may make them better able to enjoy the benefits of additional school resources, such as a better teacher, better training manuals, a library, and so on.

We turn to differences between the South and North of the country. There are dramatic differences in the programme impacts depending on whether the school is located in the South of the country or in the North (see table 1.4, or see table A.12 for standardized coefficients). In fact, if we focus on 3^{rd} grade French scores, there are no statistically significant impact of the programme in the North of the country, whereas in the South they are very large, even among boys. For example, as a result of the programme, girls in southern schools are able to increase their proportion of correct answers by almost 13 percentage points in the first follow-up, which is over 0.50 of a standard deviation. These effects are similar for other tests and persist to the end of the grade (second follow-up). When we examine all of the tests and correct the p-values for multiple testing, the impacts remain significant despite the high number of hypotheses tested. In addition, we still find no effect for fifth graders (see tables A.13 and A.14).

The South-North differences in estimates of the impact of school grants are striking. It may be the case that the types of investments made in response to the grants varied by region and took different amounts of time to manifest themselves in test scores. We examine in the subsection 1.5.3 whether there are differences between what school principals, teachers, and parents did in response to the availability of school grants in each of these areas, which could help clarify the sources of regional differences.

Table 1.3: *Programme Impacts on Grade 3 Test Scores by Baseline Ability*

Panel A: Beginning of Grade 3, First Follow-up

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

Panel B: End of Grade 3, Second Follow-up

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

[a] The index and PPVT columns are in standardized units.
Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables.
The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table 1.4: *Programme Impacts on Grade 3 Test Scores by Region*

Panel A: Beginning of Grade 3, First Follow-up

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

Panel B: End of Grade 3, Second Follow-up

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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ correspond to the Romano and Wolf's p-values from a joint test of 42 hypotheses (see text).

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

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

See text or table 1.2 for conditioning variables.

The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

In order to deepen our understanding of how the projects work, we consider three other observable student or school characteristics in our heterogeneity analyses. These variables are school size, school location and the wealth index of student's families. We distinguish between schools with small size, which have a total number of students less than the median of the distribution of that variable, and schools with large size which have a number of students greater than the median. We also distinguish between poor students and those who come from affluent contexts. We measure the wealth index and school location at the time of the first follow-up. Although it would be desirable to measure them all at baseline, doing so implies a substantial reduction in sample size, because we could not visit several schools at baseline. Furthermore, families' wealth and school location are invariant to the grants. However, we use measurements of school size at baseline since the values at follow-up surveys may reflect programme impacts.

We do not use the Romano and Wolf (2005) approach; instead, we present only individual tests. From these analyses, a mixed picture has emerged, but it mirrors the absence of effects at grade 5, other than one effect that is detected for boys attending schools with small size. These findings can be seen in tables A.15-A.20 of the appendices.

The heterogeneity analyses, when performed using school size as the interaction variable at grade 3, show an increase of females' oral scores at the first follow-up, an effect which does not persist until the second follow-up. We also observe the unexpected result that male students who attend grade 5 in small schools increase their mathematics performance at the first follow-up.

An interesting pattern has appeared when we look at the results by the levels of the wealth index. While an increase in the French and mathematics test scores is noted at the first follow-up for females from an affluent background, no effect is detected at the second follow-up for these 3^{rd} grade students. At the time of the second follow-up, all positive effects detected are for boys from rich families, and this is true for all outcomes except for the PPVT score. Thus, high socioeconomic status enhanced the impact of the programme, especially for boys on the second follow-up, who exhibit an effect of 0.36 of a standard deviation on the index score. We do not observe any effects for older students.

In rural settings, only the oral scores of female students are affected at the first follow-up. At the second follow-up, the French scores of female students in urban areas are affected. These effects are for students attending the 3^{rd} grade at the moment of the follow-up surveys. No effect is found for older students.

1.5.3 Understanding differences between South and North

We start by examining baseline test performance differences of third grade students between schools in the South and in the North. These are shown in table 1.5, panel A. Students in the southern schools perform worse on all tests (French, mathematics, and oral) than their counterparts in the North. This is particularly true for girls. When we look at the control schools in the first follow-up, documented in panel B, then the differences between the South and the North are much larger, they hold across gender groups, and they always show larger scores in the North.

Table 1.5: *South-North Differences in Test Performance, Second-Third Grades*

Panel A: Beginning of Second Grade, Baseline

	Females	Males
Percentage of correct answers: French	-0.015	0.035
	(0.027)	(0.031)
Percentage of correct answers: Mathematics	-0.068***	-0.028
	(0.024)	(0.026)
Percentage of correct answers: Oral	-0.099***	-0.078***
	(0.017)	(0.019)
Index Score	-0.357***	-0.226*
	(0.123)	(0.126)

Panel B: Beginning of Third Grade, First Follow-up

	Females	Males
Percentage of correct answers: French	-0.174***	-0.132***
	(0.026)	(0.026)
Percentage of correct answers: Mathematics	-0.145***	-0.125***
	(0.025)	(0.024)
Percentage of correct answers: Oral	-0.193***	-0.111***
	(0.022)	(0.027)
Index Score	-0.894***	-0.575***
	(0.104)	(0.125)

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

The mean test scores in the South for French, mathematics, and oral at baseline are 0.400, 0.292, and 0.134 for females and 0.461, 0.358, and 0.173 for males, respectively. At first follow-up, they are 0.426, 0.430, and 0.215 for females and 0.457, 0.486, and 0.298 for males.

Clustered standard errors in parentheses.

As mentioned above, at baseline we were only able to survey a subsample of schools. The missing schools were, as far as we can see, balanced in their treatment and control status, but they were different from the sampled schools. In fact, as we report in appendix table A.7, among control schools, missing schools are worse than the non-missing schools on a number of dimensions, as one might expect.

Therefore, it is probably safe to say that, once we look at the schools in the follow-up, which we are using to measure programme impacts, the schools in the South show much lower test results than the schools in the North. Perhaps this would indicate that they would have more room for improvement than schools in the North, and could potentially benefit from larger programme impacts. But as we mentioned above, programme impacts are larger for students with high baseline test scores.

Panel A of table 1.6 compares household characteristics of students in the South and in the North. Because of the missing schools at baseline, we take characteristics measured in the first follow-up among students in the control schools. We consider the education of the head of household and other family members, family size, number of children, a wealth index, and a proxy for the language spoken at home. A few interesting patterns emerge. Households in the southern regions are poorer but families are smaller. They have fewer children and better educated heads (and more prominently so for the families of the female students).

Panel B of table 1.6 considers the characteristics of projects being undertaken by schools with the school grant funds. This information comes from a survey conducted in treatment schools which asked principals about the project for which they got funding. We conducted two of these surveys, one at first follow-up, and one at second follow-up. We are reporting data from the second follow-up survey when, presumably, data about the project is more mature and complete.

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

Table 1.6: *South-North Differences in Grade 3 Household and Project Characteristics*

Panel A: Household Characteristics, First Follow-up

	Females	Males
Household size	-1.945***	-1.408**
	(0.594)	(0.546)
Number of children in the household	-0.743**	-0.365
	(0.366)	(0.385)
Proportion of heads of household with any education	0.118*	0.157**
	(0.070)	(0.078)
Proportion of adult females with any education	-0.033	0.072
	(0.050)	(0.062)
Wealth index	-0.829***	-0.778***
	(0.112)	(0.122)
The interview is conducted in French language	0.066	0.115*
	(0.051)	(0.059)

Panel B: Project Characteristics, Second Follow-up

	All Students
Average number of months since the project started	-7.564***
	(1.144)
Students participated in drafting of the school application	0.253***
	(0.082)
Projects with a component on Manuals	-0.095
	(0.074)
Projects with a component on Computer Materials	-0.092**
	(0.042)
Projects with a component on Teacher Training	0.162**
	(0.062)
Projects with a component on Training for the School Management Committee	0.261***
	(0.093)
Projects with a component on Courses Building	0.151***
	(0.046)
Projects with a component on General Education Improvement	0.106
	(0.100)
Project with a component on Educational Outputs Improvement	-0.015
	(0.063)
Amount spent on Principals (in 1,000,000s CFA Francs)	0.048***
	(0.014)
Amount spent on Teachers (in 1,000,000s CFA Francs)	0.039
	(0.058)
Amount spent on School Management Committee (in 1,000,000s CFA Francs)	0.087***
	(0.022)
Amount spent on Students (in 1,000,000s CFA Francs)	-0.520***
	(0.092)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Clustered standard errors in parentheses.

Some of the most remarkable differences relate to what is named as components of the project. The schools in the North were more likely to have components involving the purchase of textbooks/manuals and other educational materials, while schools in the

South were much more likely to have components related to training of teachers and of the management committee. This is also reflected in the amounts schools reported the project spent on principals, teachers, the management committee, and the students.

Therefore, we can detect clear differences in the characteristics of projects in schools in the South and the North, as stated by the principals of these schools. Schools in the South seem to be investing more in the teaching and management abilities of their human resources, while schools in the North invest more in materials. This may well be a force behind the large differences in programme impacts in these two sets of schools.

Table 1.7 reports the impact of the programme on principals' (panel A) and teachers' (panel B) behaviors. We present separate estimates of programme impacts in the South and in the North, and test whether differences in programme impacts in these two areas are equal to zero (column 3).

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

Regarding school materials and training, we see that the school grants caused an increase in books in the library in the North and an increase in the amount spent on manuals in both regions. In contrast, schools in the South spent substantially more in tutoring while both sets of schools increased spending on teacher training. All this is very much consistent with the way principals described the grant projects, as reported in table 1.6. While the point estimates reveal differences in direction in the South and North, it is difficult to be conclusive since none of the impacts are significantly different between the two regions (except expenditures on electricity and water).

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

Table 1.7: *Programme Impacts on School Characteristics by Region, First Follow-up*

Panel A: School Characteristics

	South	North	Differences
Age of the newest infrastructures	1.135	0.298	0.837
	(1.500)	(0.928)	(1.764)
The school has a library	0.201**	0.120**	0.081
	(0.086)	(0.049)	(0.099)
Number of teachers	-0.975	0.823	-1.798
	(1.085)	(0.564)	(1.223)
Number of pupils (School size)	-29.100	51.321*	-80.421
	(49.317)	(29.009)	(57.216)
Number of books in library	15.343	85.753*	-70.410
	(80.607)	(44.284)	(91.971)
Amount spent on infrastructures	40.337	53.156	-12.819
	(39.718)	(40.678)	(56.853)
Amount spent on electricity/water	-10.421	29.550*	-39.972**
	(7.867)	(15.735)	(17.592)
Amount spent on manuals	27.388**	23.019**	4.369
	(11.112)	(10.507)	(15.293)
Amount spent on tutoring	50.230*	13.512*	36.718
	(29.365)	(7.731)	(30.366)
Amount spent on teacher training	30.487**	27.856*	2.630
	(13.825)	(14.315)	(19.901)
Composition of teachers changed the year preceding the survey	-0.201**	-0.064	-0.138
	(0.086)	(0.042)	(0.096)
Percentage of female teachers	-0.031	0.012	-0.043
	(0.040)	(0.025)	(0.048)
Average age of teachers	0.273	0.315	-0.041
	(0.772)	(0.432)	(0.885)
Percentage of teachers with a Baccalaureate degree	-0.043	-0.008	-0.035
	(0.049)	(0.025)	(0.056)
Average experience of teachers	0.224	0.098	0.126
	(0.583)	(0.396)	(0.705)

Panel B: Third Grade Teacher Characteristics

	South	North	Differences
Number of minutes spent preparing lessons	3.226	1.894	1.332
	(2.941)	(2.061)	(3.591)
Number of manuals	10.475	4.990	5.486
	(7.647)	(4.980)	(9.125)
Number of measurement instruments	0.805**	-0.039	0.844**
	(0.338)	(0.208)	(0.397)
Times per day the teachers have to ask for silence	-5.060***	-0.774	-4.287**
	(1.847)	(0.699)	(1.975)
Times per day the teachers have to punish a student	0.263	-0.242	0.505
	(0.792)	(0.268)	(0.837)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Clustered standard errors in parentheses.

Finally, school grants seem to decrease teacher turnover, especially in the South. Given that teachers are likely to be the most important input in the school production function, the fact that - in the South - the programme significantly affects the amount of training they get and how likely they are to remain in the school from one year to the next is consistent with the finding of strong programme impacts on student performance in this region of the country.

Panel B of table 1.7 shows programme impacts on teacher and classroom characteristics as reported by the 3^{rd} grade teacher in the first follow-up. Contrary to what we saw above, the number of manuals are not reported by the teacher as having increased significantly either in the South or the North and measuring equipment in the South seems to have increased as a response to the programme, but not in the North (the coefficient for measuring equipment is only significant in the South).

One final and important thing to report is that, as a result of the grant, the behavior of students seems to have improved considerably in the South, but not in the North: treatment affected how often teachers have to ask for silence during the day in southern schools. This mimics what we found before when we compared 3^{rd} and 5^{th} grade teacher reports (see appendix table A.21). Student behavior improved among 3^{rd} graders but not among 5^{th} graders, which is exactly what happened in terms of test results. It is difficult to say whether test results improved because teachers were able to improve student behavior, or whether teachers who are good at teaching reading and mathematics are also able to improve behaviors in their students.

We also examined the impacts of the programme on household behaviors in the South and in the North, which is shown in table A.22. It could happen that households in the North reacted differently to the availability of school grants than those in the South, say, by investing less (substitution) in their children (which would then help explain the pattern of test results). However, this did not happen. There are no noteworthy impacts of the programme on household behaviors, and they do not seem to vary with the region of the country where households are located.

The picture relating to the differences between the South and the North that emerges from this section is mixed. The households in the South are poorer but more educated. The projects in the South tend to be more focused on training and less on information technology. However, when we look at the impact of the grant on how schools use the money, there are no obviously significant differences between the South and North. Nevertheless, the improvement in behavior in the South is remarkable and one can expect that less classroom disruption can help learning.

1.6 Conclusion

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

This paper studies the impact of a school grant programme on student performance and on potential mechanisms that could underlie the change in school performance induced by such a programme. We find strong impacts of school grants on student learning, especially on girls with high ability levels at baseline. Notably, these impacts persist over the two years of our evaluation but are not additive. We also detected positive effects of the programme for girls and boys from rich families. These impacts occur only in third grade (as opposed to later grades), and they are particularly strong in the South of the country. Our results suggest that resources distributed in a decentralized manner can have positive impacts on students.

While it is difficult to explain the grade differential in programme impacts, one conjecture is that principals focus on earlier grades because they see there the foundations for future learning. We can say a bit more, however, about the South-North differences in programme impacts, based on how we see principals spending their resources. Schools in the North emphasized manuals and other education materials whereas schools in the South emphasized human resources, namely through the training of teachers and school administrators. Our results suggest that the latter type of investments, although perhaps less visible to the local community (and therefore less preferred by say, local politicians, or even local school authorities), is likely to be more effective than the former type of investments. This result is also consistent with the idea that the main determinant of school quality is teachers, not equipment, as suggested by the most recent literature on this topic (e.g. Hanushek and Rivkin, 2006).

Chapter 2

Estimating the Causal Effect of School Size on Educational Attainment

2.1 Introduction

The literature is inconclusive as to whether small or larger schools are preferable for education systems. Recent research (e.g. Kuziemko, 2006; Schwartz et al., 2011) empirically argues the advantages of small schools, but proponents of school consolidation (Conant, 1959b; Callahan, 1962) build on the concept of economies-of-scale to justify why a larger school is better. Other empirical studies (e.g. Wyse et al., 2008) simply do not find an effect of school size on learning outcomes. Yet, most of the research on school size has been undertaken in developed countries, which deal with different challenges than those faced in low-income countries. This renders the results from existing studies on school size not necessarily generalizable to growing economies. In these latter countries, an initiative (Education for All-Fast Track Initiative,[1] EFA-FTI henceforth) has been launched to boost school enrolment. The number of children enrolled in schools in African EFA-FTI countries went up 50 percent between 2002 and 2008. In non-FTI countries, the increase was 27 percent (EFA-FTI, 2010). However, mechanisms to address the increased enrolment have not necessarily been rightly set up. Overall, teacher numbers have grown slightly less rapidly than enrolments; pupil-teacher ratios, particularly at pre-primary and primary school levels, remain high in many parts of the world and especially in Africa

[1] This is now known as the Global Partnership for Education (GPE). The initiative seeks to help low-income countries meet the education Millennium Development Goals (MDGs) and the EFA goals. A list of the goals pursue can be found at *http://www.unesco.org/new/en/education/themes/leading-the-international-agenda/education-for-all/efa-goals/*.

(Education International, 2008).

Senegal made progress expanding access to education over the last decade. The number of primary schools has increased by 57.8 percent between 2002 and 2011,[2] in part to meet the EFA objectives. However, it has been shown that quality has deteriorated (DeStefano et al., 2009) as enrolment increased. A further case is provided by national assessments (conducted by the *Programme for the Analysis of Education Systems of CONFEMEN*[3] (French acronym PASEC)) which show that scores in mathematics and French have not improved from 1996 to 2006 for either second or fifth graders (see PASEC, 2007). Not only has quality not improved as enrolment has increased, but the Senegalese government's primary response to the rising demand for education has simply been the provision of additional physical facilities. Thus, a question of interest is whether policy makers should construct more schools and keep them small or instead allow schools to have an increased number of students, while also dealing with the pedagogical issues related to both cases. This paper provides a tentative answer to the question.

We look at the impact of a primary school component, i.e. school size, on educational outcomes. The focus on this characteristic of schools stems from the fact that existing studies in the context of Senegal (e.g. DeStefano et al., 2009) never reported the role of school size as an issue that deserves thorough attention. Yet, head teachers around the world, particularly those in developing countries, face many challenges when attempting to manage large schools. In such schools, especially in the largest ones, working conditions are exceptionally harsh. While a lot is known about the causal impact of class size on students' achievement (e.g. Angrist and Lavy, 1999; Konstantopoulos, 2008; Shin and Raudenbush, 2011), little research has focused on the relation of school size to students' attainment. The existing literature examining the impact of school size typically relies only on correlation analyses (e.g. Crispin, 2011; Walberg and Walberg, 1994), and very few studies have sought to examine the causal relationship between school size and learning outcomes. The investigations into the causal gains or losses due to school size have only begun within the last two decades. While they include studies on both elementary and secondary schools, there is still not much research adopting a causal approach to studying the effects of school size on attainment in the literature. What research there is relies on randomized experiments, instrumental variables methods, or quasi-experimental

[2] Ministry of Education of Senegal, 2011.

[3] CONFEMEN stands for Conférence des Ministres de l'Education des Etats et gouvernements de la Francophonie.

strategies. Among the most credible of these studies are the experiment conducted by Bloom et al. (2010) and the research of Schwartz et al. (2011) on secondary schools, which used an instrumental variable strategy. They both suggest causal positive effects of small school size on learning achievements. In regards to elementary schools, the literature is likewise scant as to whether small size has a causal effect on learning outcomes. Assuming selection on observables and opting for the popular method of propensity score matching, Wyse et al. (2008) found no effect of large school size. However, Kuziemko (2006), who implemented a more credible strategy of instrumental variable, found a negative effect of large school size on students' achievements.

To understand the role of school size in the educational production process, we assume selection on observables. Nonetheless, our approach is differentiated by the fact that we do not match students based on their propensity score. Instead, this scalar is used to successfully weight the groups of students that we compare and make them as similar as possible outside of the fact that some are attending large schools and others small schools. Our estimations follow two main directions. Due to the lack of a clear-cut rule for defining large school in Senegal, we use the school size distribution to determine a breakpoint beyond which a school is considered large. We then compare the two types of schools using the doubly robust strategy introduced in the literature by Robins and Rotnitzky (1995). In this setting, we supplement the doubly-robust analysis with a semi-parametric and efficient analysis of quantile treatment effects as introduced by Firpo (2007). Together, linear and quantile regressions provide a comprehensive and expressive picture of the effects of school size on students' achievements in Senegal. Second, we relax our initial definition of large school size by moving up and down the threshold and by conducting extensive sensitivity analyses. These sensitivity analyses perform an important function given the complex nature of school size and the difficulty of non-randomized research: on the one hand, school size is naturally a continuous variable but has been regarded as a dichotomous treatment; on the other hand, it is hard to meet the selection on observables assumption in non-randomized research.

Our inferential framework includes a multiple inference procedure similar to that proposed in O'Brien (1984). We find robust evidence of long-term and negative effects of large school size on fourth graders and share our intuitions as to how large school size might matter. Our confidence in these results is driven by their similarities in different conditions. As already pointed out in Slate and John's (2005) review, we provide evidence that schools that are too small are not effective in fostering learning outcomes, everything else being equal. We estimate 470-500 students to be an "*ideal range*" for school size for the Senegalese context based on the effect of large school size on the overall performance of students. To the best of our knowledge, this research is the first of its kind to look at the potential effects of school size in the context of a developing economy in Sub-Saharan Africa.

The remainder of this research paper proceeds as follows: section 2.2 briefly presents the role of school size in the learning process, sections 2.3 and 2.4 respectively describe our data and the context of the research. In section 2.5, we present the methodological approach to the question, and section 2.6 lays out the results. Section 2.7 assesses the robustness of the findings and section 2.8 concludes.

2.2 School size in the learning process

The theory of change relating school size to learning achievement was developed in industrialized countries but also applies, to a certain extent, to growing economies. This theory has been supplemented by empirical findings and follows two contradictory directions.

The first strand of theoreticians and researchers support the consolidation of schools. Among them, Conant (1959b) was a pioneer in discussing the positive role of large schools, emphasizing the decreased average costs per student and the diversity of classes that a large school can offer. Callahan (1962) associated large school size with a greater opportunity to specialize while Smith and DeYoung (1988) explained that students at a small school may all be of the same demographic background. This in turn reduces the opportunity to learn from a diverse population as small school attendees are less likely to find a group of peers from which they can comfortably acquire knowledge. This theory has been supported by empirical arguments from Bradley and Taylor (1998), who find a positive effect of enrolment on exam scores, and Barnett et al. (2002), who find that large schools are more cost-effective.

Conversely, the proponents of smaller schools argue that they operate more like a community than a company. Students attending such schools are individuals and not just numbers; their academic and personal requirements are met, unlike at large educational entities where discipline problems escalate and where the odds of a student feeling isolated is much higher. Strang (1987) raised the issue of possible alienating effects due to schools with large size and stressed that the alleged specialization of teachers in larger schools mentioned by Conant (1959b) and Callahan (1962) comes at the cost of a student having many teachers and finally none of them really knowing him/her. Newman (1992) points out that large schools contribute to a lack of intellectual engagement among students. Although one of the goals in increasing school size is to offer better programs, the social needs of students may be ignored as also stated in Maxner (2005). Walberg and Walberg (1994), who are similarly proponents of small schools, underline the reinforced link between learners and community in smaller schools, a fact that they consider to be one of the advantages of such schools. From the perspective of this strand of researchers, when a school becomes larger, the overall educational process is jeopardized as bureaucracy increases. As a result of this, suppleness decreases, time spent on administrative tasks increases and students learn less because teachers teach less.

2.3 Data and attrition patterns

We use three rounds of data collected on Senegal's education system in November 2009 ($t = 0$), November 2010 ($t = 1$) and May 2011 ($t = 2$) for the purpose of a randomized evaluation of the impact of school grants on learning outcomes. The surveys cover students from the start of grade 2 to the end of grade 3 and students from the start of grade 4 to the end of grade 5.

Cognitive tests and contextual data have been gathered. The cognitive tests administration consisted of written tests of French and mathematics. An oral test was also given to students. Grade 2 written tests consisted of items covering reading comprehension and sound/written form correspondence. Grade 4 written tests focused on the functioning of the language. It included items on vocabulary, syntax, writing comprehension, spelling, grammar, and conjugation. Oral tests included topics such as reading, word recognition, non-word recognition, number of sounds, and letter recognition. As a multiple inference testing procedure, we compute a weighted average of these three outcomes by principal components analysis.[4] The use of a summary index has a range of advantages over testing individual outcomes. First, because the index represents a single measure, the likelihood of a false rejection of a hypothesis to test does not increase as additional outcomes are added to its composition; second, measuring an effect on the index provides a statistical test for whether large school size has an overall impact on the three test scores; third,

[4] Aggregations of these test scores are possible since the coefficient alpha of Cronbach is above 0.89 for both periods (2010 and 2011).

the index has more power to detect an effect (O'Brien, 1984) than individual level tests.

Contextual data includes locality-level variables, as well as head teacher- and school-level information. The survey also consists of collecting data on teachers, classrooms, students, and their households. A full list of the variables used in the analysis, together with their characteristics, is available in tables B.1 through B.3.

The sample includes 440 schools at each grade. About six students were randomly selected to take the written tests. Three of these six students passed the oral tests, and two students out of these latter three have household data. While this number of students seems small, it should be noted that the intra-class correlation coefficient is 0.52 for the French test score and 0.45 for the mathematics test score at grade 2. At grade 4, the values of the intra-class correlation coefficient are 0.53 for French and 0.57 for the mathematics test score. The values of the intra-class correlation coefficient clearly demonstrate that the driving force of the variance of learning outcomes is the number of schools rather than the number of pupils within-schools.

Attrition occurs at both the school level and student level. At the school level, one-third of the schools were not surveyed for the last data collection round. This is the case because of budgetary considerations. These schools, discarded by design, were chosen randomly. However, between the first and second rounds of data collection, no attrition occurred at school level. At student level, attrition occurred between the first two surveys, with less than 3.5 percent of the students leaving the sample at both grades. Between the second and third rounds of data collection, student-level attrition occurred not only by design because of schools discarded from the sample, but also because a group of pupils were not retrievable. The latter dropouts from the sample amount to about 11 percent at grade 2 and 12 percent at grade 4. We investigate the possibility of selective attrition at the student level using a linear probability model with the standard errors clustered at the school level. The dependent variable is a dummy equal to 1 if the student attrited, and we include our covariates in the regression models. The coefficient of the treatment indicator is insignificant in all four models estimated (see table B.5), suggesting that the results of this research are unlikely to be driven by attrition.

2.4 School size and school quality in Senegal

The average performance of students on the mathematics and French tests is very low. Second and fourth graders in 2009 correctly answered about 35 percent of questions on the French test and 30 percent on the mathematics test. On average, schools with a large size[5] have better outcomes in both topics.[6] These low outcomes are related to the overall deprived environment (see appendix table B.4). The infrastructure is quite poor, with 54 percent of schools being electrified and about 55 percent of all schools having piped water. Only 22 percent of schools possess a library. The degree of computerization is very low; there are, on average, 1.1 computers per school. When taking school size into account, our data show that large schools have more resources than small ones. The average number of computers is estimated at about three (2.62) for large schools and less than one (0.55) for small schools. Moreover, more than one out of three large schools have a library, whereas this is the case for only about one out of five smaller ones. While, in absolute numbers, large schools have more manuals than small schools do, the apparent advantage is diluted by school size and results in no relative primacy. In fact, only eight French manuals and three mathematics manuals are accessible per ten students which highlights again the deficiencies of the learning context.

It is also seen that small schools are characterized by relatively small class sizes compared to larger schools. Depending on the grade, the average class size of small schools fluctuates between 31 and 34 students; in larger schools, class size varies from 49 to 58 students. This observation tends to highlight the demographic pressure of geographic areas where large schools are located. Indeed, the survey shows that small schools are located in areas with fewer than 50,000 inhabitants, while larger schools are found in areas where the population is more than six times larger. This difference in population is probably related to the fact that 86 percent of small schools are rural whereas around 42 percent of large schools are located in urban areas. The isolation of schools in Senegal is evidenced by the average distance of 336 kilometres that separate them from the national capital (Dakar), the center of all economic and development activities of the country. Even the closest urban centres are relatively far away, located 19 kilometres on average from schools; for some schools, the distance to the closest urban centre may even be as much as 225 kilometres. This analysis, repeated by type of schools (large or small), reinforces the finding of population pressure stated above. Large schools are much closer to urban centres or the national capital, the most populous places in the country.

[5]The terms "schools with a large size" and "large schools" are used interchangeably to improve readability. The same is true for "schools with a small size" and "small schools."

[6]School size is studied in this paper as a dichotomous treatment variable. We provide reasons for this later in this paper and discuss how we define small and large schools.

Nearly three-quarters of schools are close to a health centre. Taking into account the size of the school shows that larger schools are closer to safer environments. Indeed, over 95 percent of these schools have a health centre nearby, 74 percent are electrified (fully or partially), and about 80 percent have running water. Regarding schools with a small size, 62 percent have a health centre nearby. Just 47 percent are fully or partially electrified, and the same proportion (47 percent) has running water. Moreover, while the average age of school facilities is 37 years for large schools, it is 21 years for small schools. This tends to confirm that urban centres benefit from the construction of infrastructure sooner than rural ones because they have to match the strong and fast-growing demand for education.

School size is highly correlated with the number of teachers and classrooms. On average, large schools have 15 teachers and 12 classrooms, but small schools only possess seven teachers and six classrooms. On average, staff academic qualifications are relatively low. About 70 percent of the head teachers hold a Baccalaureate degree or the equivalent professional qualification.[7] The rest (30 percent) possesses the BFEM.[8] Managers of large schools are older and more experienced than their colleagues, the leaders of small schools. It also appears that there is a greater presence of women in the management of large schools. Indeed, they represent 4.1 percent of school principals of large schools, compared to just 1.4 percent of heads of small schools. However, the participation of parents or the schools' management committees in the decision-making process is more or less the same between the two groups of schools.

2.5 Empirical approach

2.5.1 School size as a continuous treatment

A key empirical challenge in identifying the causal impact of school size on learning outcomes is that school size is not randomly determined;[9] school size is instead determined by many factors. It can be affected by enrolment limitations or encouragement in relation to the push for Education for All. Schools may attract numerous students because of the

[7] This is the Elementary School Teaching Diploma (French acronym BSEN, Brevet Supérieur d'Etude Normale).

[8] BFEM is Brevet de Fin d'Etudes Moyennes. It is a secondary school certificate obtained after ten years of schooling.

[9] The issue of school choice that generally plagues the school quality literature is discussed in Altonji et al. (2005) in the context of selection into Catholic and public schools.

school-age population combined with the number of schools in its catchment area. It may also attract students if the teachers are considered to be better than others or if it has infrastructure and facilities that are in better working conditions.

At least two methods can be used to identify the causal effects of school size on learning outcomes if one looks at school size as a continuous treatment: the instrumental variable regression and the dose-response model. In the context of this research, it is inappropriate to argue that any of the available background, family, class, or school variables determines school size but not learning outcomes. In the absence of an appropriate instrument, a frequently used way to circumvent the endogeneity issue is to assume that, conditional on a set of observable variables, the treatment (school size) can be considered to be as good as randomly assigned. Our methodology is built upon this assumption.

Following Imbens (2000), Hirano and Imbens (2004) propose an extension to the propensity score method that applies to continuous treatments.[10] They define a generalization of the binary treatment propensity score which has many of the appealing properties of the binary treatment propensity score. For the estimated effects to have a causal interpretation, it is essential that, given the generalized propensity score, assignment to any value of school size is independent of the potential outcomes. However, when implemented, the generalized propensity score fails to balance the data, preventing the drawing of any causal inference based on the dose-response method. This leaves room for addressing the research question using alternative strategies.

2.5.2 School size as a dichotomous treatment

This method is a second best substitute to looking at school size as a continuous treatment. Here, we distinguish between large and small school size. Because school size is not a random variable, taking the mean difference of test scores between the group of students attending large schools and the one of learners attending small schools will not yield the net impact of school size on students' attainment. To see why this is the case, we define $T_i = 1$ if student i attends a large school and $T_i = 0$ if he or she attends a small learning entity. If $Y_i(T_i) = (Y_i(0), Y_i(1))$ denotes the pair of potential outcomes student i attains if he or she is exposed to the treatment or the control group, then we have:

$$\begin{aligned}\mathbb{E}(Y|T=1) - \mathbb{E}(Y|T=0) &= \mathbb{E}(Y(1)|T=1) - \mathbb{E}(Y(0)|T=0) \\ &= \mathbb{E}(Y(1)|T=1) - \mathbb{E}(Y(0)|T=1) + \mathbb{E}(Y(0)|T=1) - \mathbb{E}(Y(0)|T=0)\end{aligned} \tag{2.1}$$

[10]Also see Imai and Van Dyk (2004) for causal inference in general treatment regimes.

$\mathbb{E}(Y(1)|T=1) - \mathbb{E}(Y(0)|T=1)$ is the average effect of school size on test scores of pupils attending large schools and $\mathbb{E}(Y(0)|T=1) - \mathbb{E}(Y(0)|T=0)$ is the selection bias. This bias originates from the fact that the average outcome of students who attended schools with large size would have been different than that of students attending schools with small size had they attended schools with small size. This is the case because these two populations are not identical. For instance, our data reveals that 60 percent of urban fourth graders go to large schools. Only 16 percent of rural fourth graders attend large schools. The location of the school (as well as other variables) is then an important confounding factor, and not controlling for it will lead to an estimation fallaciously attributed to the treatment status.

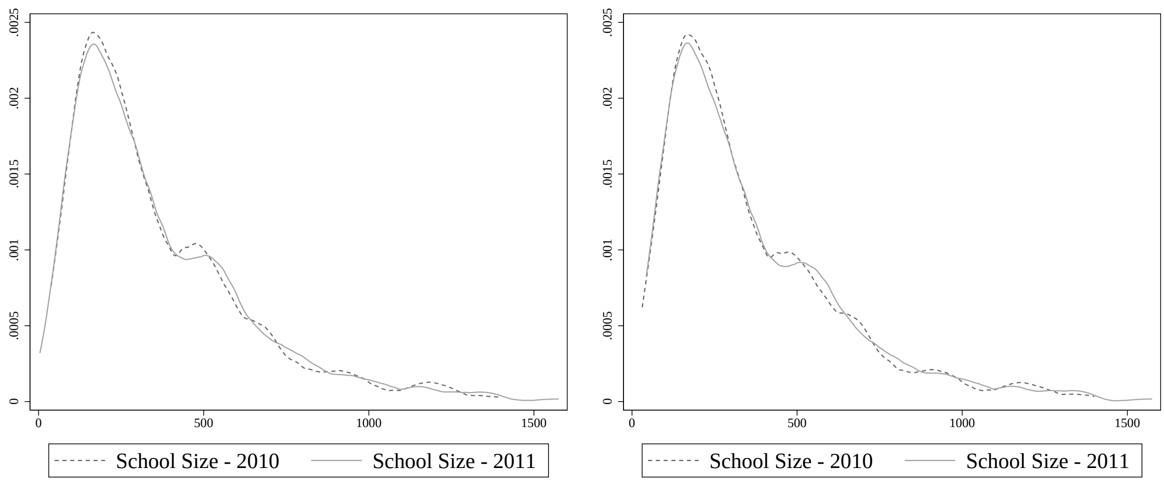
We assume that, conditional on a set X of observable variables, the treatment T is as good as randomly assigned, that is $(Y_i(0), Y_i(1)) \perp\!\!\!\perp T_i|X_i$. This is a plausible assumption since we control as much as possible for confounding factors suggested in the literature of education production functions or by the theory related to school choice. Also, because we include students' ability at the start of the school year in the estimated models, we can think of the models as growth models (or value-added models). Hanushek (1979) states that controlling for prior ability alleviates problems of omitting prior inputs of schools and families because the initial achievement level already includes the missing information. Also, research evidence (e.g. Kane and Staiger, 2008) suggests that the resulting bias in value-added models is reasonably small. Thus, even if the conjecture of selection on observables is not fully correct, any bias that exists will not prevent the estimated models from detecting important aspects of the effects of interest.

2.5.2.1 Construction of the treatment status

We dichotomize the continuous variable containing school size to form the groups of school that will be compared. For this to be reasonable, groups within each arm of the treatment variable need to be "*homogeneous*" in terms of treatment status. In other words, schools called "*treatment*" must be similar and very different from comparison ones. We achieve this through hierarchical cluster analysis, which has the advantage of proposing a cut-off point based on the distribution of school size in the absence of a formal definition of a large school. Amongst the most popular algorithms for hierarchical clustering, we use Ward's linkage which is perceived to be very efficient in that it uses an analysis of variance to evaluate the distances between clusters. This form of hierarchical clustering minimizes the within-clusters variance while simultaneously maximizing the between-clusters variance.

We check the stability of school size between the school years 2009-2010 and 2010-2011. As can be seen on figure 2.1, there is no change in this variable over time. The fact that the distribution of school size did not statistically change[11] indicates the stable intensity of the treatment students are exposed to and offers the possibility to look at the effect of the same large school size in the short run and after 18 months of exposure. For the remainder of this paper, we distinguish between the short-term effects that could appear after a year of exposure and the long-term effects that are detectable (if they exist) after 18 months of exposure to the same large school size.

Figure 2.1: *School Size Density for Grade 2 (left side) and Grade 4 (right side) at School Years 2009/2010 and 2010/2011*



2.5.2.2 Outcomes of interest

Throughout this paper, we model three types of outcomes that measure only cognitive skills, though school size could impact non-cognitive outcomes as well (e.g. disruption, truancy, teacher quality, school attendance, dropout rates, parent involvement). In addition, we model a summary index of the learning outcomes. Even though school size could have affected non-cognitive outcomes, we focus on cognitive aspects of students' development as our primary interest. Cognitive abilities are rewarded in the labor market, as emphasized in Hanushek (1972), and they are also good predictors of students'

[11] We perform a Wilcoxon test for the equality of the distribution of school size at 2009 to the distribution of school size at 2010. The tests cannot reject the equality of the distributions as the p-value is 0.65 for grade 2 data and 0.68 for grade 4 data.

subsequent progress.[12]

2.5.2.3 Conditioning variables

There is a debate in the literature over which variables to include in the matrix X. Rubin and Thomas (1996), Heckman et al. (1998), and Glazerman et al. (2003) argued that it is crucial to include in the matching procedure all covariates perceived to be correlated with both the treatment assignment and the outcome so that the selection on observables assumption, also referred to as the unconfoundedness assumption, is likely to be met. While methods that use a limited set of covariates perform poorly (Shadish et al., 2008), in small samples one should primarily be concerned with variables supposed to be associated with the outcome as there is a higher increase in variance due to the inclusion of variables orthogonal to the outcome but forcibly-connected to treatment assignment (Brookhart et al., 2006). We follow this latter direction in this research. All covariates included in this paper, together with their means and standard deviations, are presented in tables B.1-B.3, as we already mentioned. These covariates include time-invariant variables, as well as variables that could change in a way unrelated to school-level shock. We discuss in the robustness checks section how our estimates behave in relation to a particular set of covariates.

2.5.2.4 Parameters of interest

We are interested in estimating the average treatment effect of large school size on a randomly chosen student (average treatment effect, ATE), as well as the impact of large school size on students attending large schools (average treatment effect on the treated, ATET). We also want to go beyond the mean and provide some insights on the distributional impact of school size on learning achievements. For this, we estimate, in the first place, the (unconditional) quantile treatment effects.

2.5.2.5 Regressions analyses[13]

As an entry point for our discussions, we compute a naïve difference of average learning outcomes between large and small schools. This is operationalized by running a regression

[12] Heckman et al. (2006) discuss the effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior, and Heckman and Rubinstein (2001) discuss the importance of non-cognitive skills.

[13] In this paper, we prefer regression to matching for different reasons: the need to construct standard errors accounting for clustering, the failure of bootstrap in the context of matching (Abadie and Imbens, 2008) and the practical difficulty of finding good matches in the presence of several continuous variables. We discuss in subsection 2.6.3 similarities between matching and regression.

of each outcome to the treatment indicator and the intercept. Second, we estimate conditional models (on the whole sample and on the common support of the data[14]) which take the form:

$$Y_{ist}^{k} = \alpha_t^k + \Delta_t^k T_s + X_{is}\delta_t^k + \varepsilon_{ist}^k \tag{2.2}$$

where Y_{ist}^k is the standardized score in test k, for student i in school s at period t (1 or 2), T_s is a treatment indicator, X_{is} are conditioning variables measured in 2009 (period $t = 0$) and ε_{ist}^k is a disturbance term such that $\mathbb{E}(\varepsilon_{ist}^k) = 0$. The matrix X_{is} will first contain data on school quality plus some basic student characteristics (age, gender, prior ability). In a subsequent stage, we include more details on the students' personal characteristics and families' contexts. Since only one-third of the sample has been randomly picked for the rest of the survey, the latter estimations are based on a reduced sample size. Nonetheless, there is not much power loss due to collecting data on a reduced number of students within each school as the sample is clustered at the school level.

For simplicity, we discard the superscript k from the following notations. We also use $T_i = 1$ (instead of $T_s = 1$) to indicate that student i attends a school with large size or X_i instead of X_{is} to refer to the pre-treatment matrix of student i in school s. All standard errors derived from these regressions are robust to heteroskedasticity and clustered at the school level.

2.5.2.6 Doubly robust regression

Identification through doubly robust methods involves two steps: the weighting based on the propensity score and a covariance adjustment. This method is discussed in the literature by Robins and Rotnitzky (1995), Robins, Rotnitzky and Zhao (1995), and Van Der Laan and Robins (2003). Combining regression adjustment and inverse propensity score weighting accomplishes some robustness to misspecification of the models by both removing the correlation between the omitted covariates and by reducing the correlation between the omitted and included variables, as stated in Imbens and Wooldridge (2009). The presentation of this method that we adopt here is, to a large extent, borrowed from these authors.

The idea of inverse propensity weighting is the same as for inverse probability weight-

[14] In this second case, we estimate the conditional expectation function $\mathbb{E}(Y_{ist}^k | \widehat{p}(T_s = 1|X_i) \in S_0 \cap S_1) = \alpha_t^k + \Delta_t^k T_s + X_{is}\delta_t^k$ where $\widehat{p}(T_s = 1|X_i)$ is the predicted probability of being treated and S_0 and S_1 are the intervals of this estimated probability for non-treated and treated students, respectively.

ing in survey research (see Horvitz and Thompson, 1952). This method down-weights students that are overrepresented in small or large schools and up-weights those that are underrepresented in any of the groups.

Let us denote $p(T_i = 1|X_i) = p(X_i, \xi)$ the probability of attending a large school where ξ is a parameter in the probability model. $p(X_i, \xi)$ is our propensity score. The weight scheme for the estimation of the average treatment effect is such that $1/p(X_i, \xi)$ is used to weight each observation in the treatment group and $1/(1 - p(X_i, \xi))$ is used as weight for the comparison group members. From the relation $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$, we obtain $T_i Y_i = T_i Y_i(1)$ and can write:

$$\begin{aligned}
\mathbb{E}\left\{\frac{T_i Y_i}{p(X_i, \xi)}\right\} &= \mathbb{E}\left\{\frac{T_i Y_i(1)}{p(X_i, \xi)}\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\frac{T_i Y_i(1)}{p(X_i, \xi)}|X_i\right]\right\} \\
&= \mathbb{E}\left\{\frac{\mathbb{E}(T_i|X_i)\mathbb{E}(Y_i(1)|X_i)}{p(X_i, \xi)}\right\} \\
&= \mathbb{E}\left\{\frac{p(X_i, \xi)\mathbb{E}(Y_i(1)|X_i)}{p(X_i, \xi)}\right\} \\
&= \mathbb{E}\left\{\mathbb{E}(Y_i(1)|X_i)\right\} \\
&= \mathbb{E}\left\{Y_i(1)\right\}
\end{aligned} \tag{2.3}$$

The second and final equalities are due to iterated expectations. The third one follows by quasi-randomization. Similarly, from $T_i Y_i = (1 - T_i) Y_i(0)$, the following holds:

$$\mathbb{E}\left\{\frac{(1 - T_i) Y_i}{1 - p(X_i, \xi)}\right\} = \mathbb{E}\{Y_i(0)\} \tag{2.4}$$

Weighting the sample of students attending large schools by the inverse of the propensity score recovers the expectation of the unconditional outcome under treatment. Likewise, weighting the sample of students attending small schools by the inverse of the probability of non-treatment recovers the expectation of the unconditional outcome in the absence of treatment. The interpretation of the propensity score weighting approach is that it creates a pseudo school population in which there is no confounding variable so that the weighted averages reflect true means in the real world. Given equations 2.3 and 2.4, the inverse propensity score weighted estimator of the average treatment effect can be written as:

$$\triangle = \mathbb{E}\left[\frac{T_i Y_i}{p(X_i, \xi)} - \frac{(1 - T_i) Y_i}{1 - p(X_i, \xi)}\right] \tag{2.5}$$

Surprisingly, Hirano et al. (2003) showed that the use of an estimated propensity score leads, asymptotically, to a more efficient estimator than the use of the true propensity score. Thus, one can estimate the average treatment effect using the following equalities:

$$\widehat{\triangle} = n^{-1}\sum_{i=1}^{n}\frac{T_iY_i}{p(X_i,\hat{\xi})} - n^{-1}\sum_{i=1}^{n}\frac{(1-T_i)Y_i}{1-p(X_i,\hat{\xi})} \tag{2.6}$$

If the average treatment effect of school size on large school attendees is of interest, a major modification is required in the weight scheme. Instead of being equal to $1/p(X_i,\hat{\xi})$ for the treated and $1/(1-p(X_i,\hat{\xi}))$ for the controls, the weights are 1 for individuals of the treatment group and $p(X_i,\hat{\xi})/(1-p(X_i,\hat{\xi}))$ for those of the control group. In this design, the control group is weighted to represent the average outcome that the exposed group would have exhibited in the absence of treatment.

In the first step of the doubly robust method, we calculate ξ by maximum likelihood, obtain the estimated propensity scores as $\hat{p}(T_i = 1|X_i) = p(X_i,\hat{\xi})$ and construct the weights as above.

The second step involved in this method is the covariance adjustment. Suppose we model the two regression functions for the potential outcomes as $\mu_T(x) = E(Y_i(T)|X_i = x) = \alpha_T + \beta_T'(x-\bar{X})$ where X_i is a matrix of covariates for student i, $\bar{X}$ is the average of the X_i, x is a realization of X_i on our sample, $T = 0$ for the control group and $T = 1$ for the treatment group. By construction, the average treatment effect conditional on $X = x$ is $\mu_1(x) - \mu_0(x)$. We use linear regression, where we weight the objective function by the inverse probability of treatment or non-treatment. Explicitly, to estimate (α_0,β_0) and (α_1,β_1), we solve the following weighted least squares problems:

$$\underset{\alpha_0,\beta_0}{min}\sum_{i:W_i=0}\frac{(Y_i-\alpha_0-\beta_0'(X_i-\bar{X}))^2}{1-p(X_i,\hat{\xi})} \quad ; \quad \underset{\alpha_1,\beta_1}{min}\sum_{i:W_i=1}\frac{(Y_i-\alpha_1-\beta_1'(X_i-\bar{X}))^2}{p(X_i,\hat{\xi})} \tag{2.7}$$

The doubly robust estimate of the average treatment effect, based on the predicted propensity score and potential outcomes, is $\widehat{\Delta} = \hat{\alpha_1} - \hat{\alpha_0}$ and takes the following form:

$$\widehat{\Delta} = \mathbb{E}\left[\frac{T_iY_i}{p(X_i,\hat{\xi})} - \frac{T_i-p(X_i,\hat{\xi})}{p(X_i,\hat{\xi})}\mu_1(X_i,\hat{\beta})\right] - \mathbb{E}\left[\frac{(1-T_i)Y_i}{1-p(X_i,\hat{\xi})} + \frac{T_i-p(X_i,\hat{\xi})}{1-p(X_i,\hat{\xi})}\mu_0(X_i,\hat{\beta})\right] \tag{2.8}$$

As in the case of the inverse propensity weighted estimator, modifying the weight scheme to 1 for treatment students and $p(X_i,\hat{\xi})/(1-p(X_i,\hat{\xi}))$ for those of the comparison group allows for the computation of the doubly robust estimate of the average treatment effect on the treated. A simple reorganization of formula 2.8 suggests that the doubly robust estimator equals the inverse propensity score weighted estimator augmented by a second term. See Tsiatis (2006) for a proof of the property of the doubly robust estimator.

Like previously, this strategy needs to deal with the hierarchical structure of the data. The regressions analyses specified above easily take into account the multilevel form of the data by clustering the standard errors of the estimates at the school level. However, propensity score models have been discussed and constructed in the literature when units are randomly sampled; only a few examples deal with propensity scores in hierarchical settings (Lingle, 2009; Su and Cortina, 2009), as is the case for this research.[15] We deal with heteroskedasticity of standard errors and clustering at both stages of this method, though accounting for this only in the second step is sufficient to properly compute the standard errors of the point estimates. Indeed, the correctness of the standard errors estimated at the first step does not affect the predicted probability of being treated.

Because students have been followed-up with twice, we produce two estimations of the ATE using equation 2.8 and, likewise, we produce two estimations of the ATET. Again, we assume that students who share the same observable dimension X_i but who attended schools characterized by different values of T_i do not differ, on average, in the unobserved dimension ε_i. This means that $\mathbb{E}(\varepsilon_i|X_i, T_i) = \mathbb{E}(\varepsilon_i|X_i)$. The matrix X_i is used to control for remaining imbalance (if any) and improve the precision of the estimates.

2.5.2.7 Quantile treatment effects and quantile treatment effects on the treated

Quantile treatment effects are defined and introduced in the literature by Doksum (1974) and Lehman (1974). Let q be a real in [0-1]. Symbolically, for any fixed quantile q, the *q-th* quantile treatment effect is defined as $\Delta_q = \tau_q = \tau_{1,q} - \tau_{0,q} = F^{-1}_{Y(1)}(q) - F^{-1}_{Y(0)}(q)$, where F^{-1} denotes the inverse of the cumulative distribution function F of Y and for $j \in \{0; 1\}, \tau_{j,q} = inf_q \Pr[Y(j) \leqslant \tau] \geqslant q$. Thus, the quantile treatment effect is defined as the horizontal distance between two marginal distributions, for any fixed quantile. This is different from the quantile of the unit level effect $\tilde{\tau}_q = F^{-1}_{Y(1)-Y(0)}(q)$. Generally, these two quantities are different, except in the particular case of a perfect rank correlation between the distributions of the potential outcomes. The first estimator has received more attention in the literature.[16] We are interested in the quantile treatment effect and

[15] Li et al. (2013) propose a study of propensity score weighting with multilevel data.

[16] Imbens and Wooldridge (2009) provide substantial and statistical reasons for why this estimator is preferred to the quantile of the unit level effect.

in the quantile treatment effect on the treated which is $\Delta_q^{TT} = \tau_{q|T=1} = \tau_{1,q|T=1} - \tau_{0,q|T=1} = F^{-1}_{Y(1)|T=1}(q) - F^{-1}_{Y(0)|T=1}(q)$, where, for $j \in \{0;1\}$, $\tau_{j,q} = inf_q \Pr\left[Y(j) \leqslant \tau | T=1\right] \geqslant q$.

Identification of the distributional effects of a given treatment requires strong ignorability (unconfoundedness and common support) and uniqueness of quantiles.[17] Two competitive estimators are the conditional quantile treatment effect and the unconditional quantile treatment effect. The unconditional quantile treatment effect has the advantage over the conditional quantile treatment effect in that its meaning does not change with the set of covariates used. Not only does it summarize the effects of a treatment for the whole population, but it is also usually paid the most attention in policy evaluations since the results can be conveyed and summarized in a straightforward way (Frölich and Melly, 2013). Moreover, unconditional quantile treatment effects can be estimated at the $\sqrt{n}$ convergence rate without any parametric restriction, which is impossible for conditional quantile treatment effects, unless all covariates are discrete (Frölich and Melly, 2010, 2013). Nevertheless, covariates are needed to remove selection bias (Firpo, 2007). In this setting, covariates play the same role as when estimating average treatment effects under the assumption of selection on observables. The method implemented is suggested by Firpo (2007), who proposed a two-step approach. First, the propensity score that has been estimated in the previous analyses is supplied here since, as we will see later, the weights constructed based upon it led to a very good balance of the covariates. Second, we assume the relation $\mathbb{Q}^q(Y|T) = \alpha^q + \Delta^q T$ and estimate the quantile treatment effects and the quantile treatment effects on the treated by solving the following optimization program:

$$(\widehat{\alpha}^q, \widehat{\Delta}^q) = \underset{\alpha,\Delta}{argmin} \sum_i w_{i,j} * \rho_q(Y_i - \alpha - \Delta T_i) \tag{2.9}$$

where $w_{i,j}$ is the weight of student i in group $j \in \{0;1\}$, $\rho_q(u) = u * \{\tau - \mathbb{1}(u<0)\}$ is the check or loss function with $\mathbb{1}(.)$ the usual indicator function. The solution to equation 2.9 is the traditional inverse probability weighting estimator. The weights are built according to the estimator we are interested in. For estimating the quantile treatment effects, the weights are $w_{i,1} = \frac{T_i}{N*\widehat{p}(T_i=1|X_i)}$ for students of the group of large schools and $w_{i,0} = \frac{1-T_i}{N*(1-\widehat{p}(T_i=1|X_i))}$ for students of the group of small schools. For estimating the quantile treatment effects on the treated, the weights are $w_{i,1} = \frac{T_i}{\sum_{k=1}^{N} T_k}$ for students of the group

[17] See Firpo (2007) for further details.

of large schools and $w_{i,0} = \frac{\widehat{p}(T_i=1|X_i)}{1-\widehat{p}(T_i=1|X_i)} * \frac{1-T_i}{\sum_{k=1}^{N} T_k}$ for students of the group of small schools.[18] T_i is the treatment status of the school the student attends, whereas the matrix X_i is his or her set of covariates.

2.6 Results and discussions

This section presents and discusses the results obtained from the empirical strategy implemented. First, we present the results from the cluster analyses that determine the treatment status of the schools. Second, we present the results from the various regressions. Throughout this section, the analyzed learning outcomes have been demeaned and converted to effect sizes by dividing each demeaned outcome by its standard deviation.

2.6.1 Cluster analyses

The cluster analyses have been run on separate data for grade 2 and grade 4. The data reveal several influential points in the distribution of school size. Indeed, starting with 440 schools (for each grade), we finally discard 25 outliers schools (identified by the boxplots in figure B.1) with a size greater than 869 students.[19]

The cluster analysis proposes a cut-off point of 440 students to distinguish between large schools and small schools.[20] From our grade 2 data, 115 schools have 440 students or more and can thus be considered large schools, while 300 schools have fewer than 440 students and can be viewed as small schools. At grade 4, the hierarchical cluster method proposes a partition of 114 large schools and 301 small schools. At both grades, the large schools' size has an upper bound of 869 students. It should be noted that 112 large schools and 283 small schools encompass our grade 2 and grade 4 classrooms; the

[18] Equations 2.3 (for the treatment group) and 2.4 (for the comparison group) are true not only for Y but also for any measurable function $g(Y)$. Under the assumptions of strong ignorability and uniqueness of quantiles, Firpo (2007) proves that the quantile treatment effect and the quantile treatment effect on the treated become estimable from the data. Quantiles of the potential outcomes distributions are indeed implicit functions of the outcome variable, the treatment and the covariates. Using equation 2.3 and $g(Y) = g_1(Y) = \mathbb{1}\{Y \leqslant \tau_{1,q}\}$ for the treatment group and using equation 2.4 and $g(Y) = g_0(Y) = \mathbb{1}\{Y \leqslant \tau_{0,q}\}$ for the comparison group, where $\tau_{j,q} = argmin_{\alpha,\Delta} \sum_i w_{i,j} * \rho_q(Y_i - \alpha - \Delta T_i)$ for $j \in \{0;1\}$, $\rho_q(.)$ is the check or loss function, $\mathbb{1}(.)$ is the indicator function, and the weights $w_{i,j}$ are expressed as usually for each group, we can obtain the moment conditions for the quantile functions and show that $\mathbb{E}\left\{\frac{Tg_1[Y(1)]}{p(T=1|X)}\right\} = \mathbb{E}\left\{g_1\left[Y(1)\right]\right\}$ and $\mathbb{E}\left\{\frac{(1-T)g_0[Y(0)]}{1-p(T=1|X)}\right\} = \mathbb{E}\left\{g_0\left[Y(0)\right]\right\}$. The quantile treatment effect on the treated is estimated by posing $g(Y) = g_1(Y) = \mathbb{1}\{Y \leqslant \tau_{1,q|T=1}\}$ for the treatment group and $g(Y) = g_0(Y) = \mathbb{1}\{Y \leqslant \tau_{0,q|T=1}\}$ for the comparison group.

[19] The sensitivity of the estimates to this exclusion is checked in section 2.7.4.

[20] The cluster analysis has been run on the whole set of schools, including outliers schools on the one hand and excluding outliers schools on the other hand. Both analyses led to the same results.

remaining schools have either grade 2 and 4. This definition of large and small school size is the one we adopt until the robustness of the results is shown or refuted.

2.6.2 Basic regressions

We present findings from standard regressions in the appendix table B.6 for second graders and the appendix table B.7 for grade 4. At all grades, the naïve models suggest positive and significant impacts of large school size in the short run.

After 18 months of exposure, the effects all vanish at both grades 2 and 4, except for the oral test at grade 2. These findings are concordant with the previous description of school quality in Senegal. Large schools tend to be close to urban zones, are in better working conditions, and have more financial resources and learning materials than small ones. However, these gross estimates are misleading, as the effects of confounders have not been separated out. A step toward this direction involves confounding covariates in a regression model.

At grade 2, and irrespective of the set of covariates we use, the models suggest no effect after 12 or 18 months. At grade 4, we have detected negative impacts of large school size on long-term outcomes (the oral test score and the index). While this is true for all set of covariates, it should be noted that inclusion of more covariates yields effects that are greater in absolute value and more significant. These regressions that control for confounding factors certainly represent an improved strategy, but they do not offer a flawless solution to the causal inference that is intended to be made. Indeed, the selection on observed covariates assumption does not guarantee that a regression of learning outcomes on the confounding covariates and a treatment indicator is the best modelling approach for estimating treatment effects. This is the case because lack of overlap and imbalance are threats to the validity of the comparison between schools with large size and schools with small ones. Using ordinary least squares, we cannot make any causal inference in the regions of the covariates where there are no large or small school.

In a subsequent stage, we constrain the estimation on the common support of the data, while still using ordinary least squares. Running ordinary least squares on the common support produces little gain as the method suggests no effect at grade 2, whether in the short run or not, independently of the set of controls. At grade 4, however, there are negative and longer-term effects on the oral test and on the index. A negative effect is also observed in the short run with the second (full) set of conditioning variables.

2.6.3 Doubly robust estimation

The regression models estimated in the previous subsection do not recover our parameters of interest (average effect of large school size and average effect of large school on students attending such learning centres). This is the case because regression is a form of weighted matching estimator that put weights proportional to the conditional variance of treatment whereas matching weights are proportional to the probability of treatment at each value of the covariates (Angrist, 1998). Yet, both methods require the conditional independence assumption to hold so that the estimates have a causal interpretation. Thereby, the main concern in this research design is much related to the likelihood of the stated assumption than the form of technique used to apprehend the effects of school size.

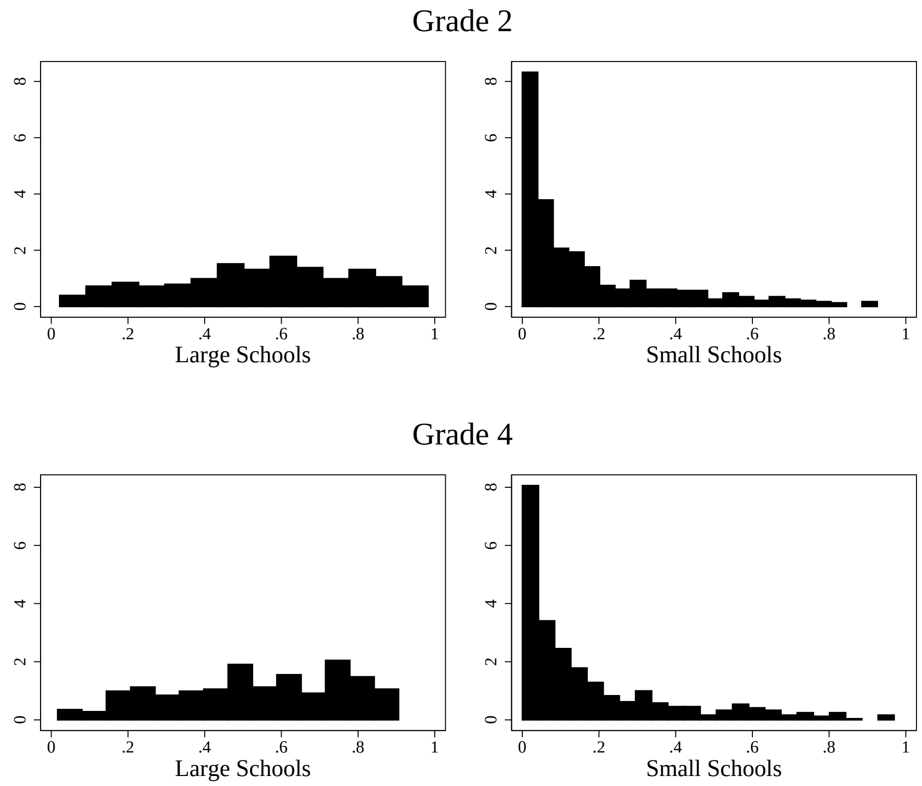
To properly recover our parameters of interest (average treatment effect and average treatment effect on the treated), two types of weights are borrowed from the literature of causal inference. The use of these weights constructs comparable groups of schools by re-building the distribution of the control group if the aim is to estimate the average treatment effect on the treated and the distributions of both groups if the objective is to estimate the average treatment effect. Given this weighted sample and selection on observables, attendance of any type of school is random.

2.6.3.1 Propensity score estimation

The estimation of the propensity score has been conducted using a standard probit model on the two sets (incomplete and full) of covariates. The model based on the full set of covariates can be found in the appendix table B.8. The appendix table B.15 presents the propensity score model based on the reduced set of covariates. Balance of the covariates is not fully achieved with the incomplete set (first set) of controls,[21] even if the second step of the doubly robust regression can correct for it. Figure 2.2 below provides an overview of the distributions of the estimated propensity scores for grades 2 and 4. Distinct distributions are built for students attending large schools (440-869 pupils) and students attending small schools (fewer than 440 students) so that it is possible to visually check how the groups in comparison overlap. The two groups are noticeably different, and the support of the estimated propensity scores is almost the complete interval [0-1]. The quality of the overlap is much better at grade 2, as at grade 4, the observed maximum propensity score is away from one.

[21] We discuss this in the next subsection.

Figure 2.2: *Propensity Score Histograms*



The graphics of the propensity scores indicate that the probability of attending a school with large size is high for a majority of students going to large schools. Conversely, for the majority of students attending a school with small size, the propensity score is quite small, even if there is some students with high probability to attend a school with large size.

2.6.3.2 Balancing checks

We use both joint and independent balance tests to assess the degree of selection in the samples. The independent balance tests are performed by running a regression of each covariate on the treatment indicator plus the intercept, whereas the joint balance tests consist of standard probit regressions of the treatment indicator on the whole set of covariates plus the intercept. The t-statistics resulting from the independent balance tests, where we compare small schools with large ones, can be found in tables B.13 or B.20.[22] The comparison has been made on a set of 41 covariates. Based on independent tests,

[22]The statistics in the balance checks tables may be somewhat different than some statistics we already presented in section 2.3 where we provided a descriptive comparison of small and large schools. The reason for this is that the sample in the previous analysis is made of schools that either have grades 2 and 4 or only have one of grade 2 or 4, while this analysis is conducted by grade.

24 variables in grade 2 and 20 variables in grade 4 were not balanced before weighting. We also report normalized differences (see table B.14) proposed by Imbens and Rubin (1997). These are defined as the between-treatment differences in means, scaled by the average of the two within-treatment standard deviations. This statistic, which is invariant to sample size, is a more reliable way of assessing covariate balance. A normalized difference of 0.25 or less (in absolute value) indicates satisfactory balance. The degree of selection reflected by this criterion is lower at both grades.

Weighting using a propensity score predicted with the whole set of covariates fully balances the samples for both grades and for both estimators if one considers the independent balance tests (see table B.13). The joint balance tests yield p-values of almost one for both estimators and both grades when all covariates are used in the estimation of the propensity score. Tables B.9-B.12 show the results for these joint balance tests. It is worth mentioning that the propensity score predicted with the reduced set of covariates is less successful in achieving balance, but it does improve the balance, though not all variables were included in its computation: depending on the estimator, between 34 and 37 variables at grade 2 and 38 and 39 variables at grade 4 have been balanced. This is visible in the appendix table B.20. The use of the reduced set to estimate the propensity score leads to p-values of the joint balance tests that are only slightly improved at grade 2 but substantially increased at grade 4 (see tables B.16-B.19). Such results denote the effective operation of the probability of treatment as a balancing score, especially when its prediction is based on our whole set of covariates.

2.6.3.3 Inverse probability weighting and regression adjustment

Using inverse probability weighting, the estimator is sensitive to extreme values of the propensity score which results in singularly large weights. The estimation of the ATE is sensitive to values close to the bounds of the probability interval, whereas the estimation of the ATET is sensitive to values close to zero only. As a rule of thumb, Crump et al. (2009) suggest the use of a trimming level of 0.10, meaning that all observations such that $0.10 \leq \widehat{p}(X_i) \leq 0.90$ should be excluded from the analyses when aim is to estimate the average treatment effect. We follow this rule when the purpose is to estimate the average effect of large schools. We discard all observations with a propensity score smaller than 0.10 when focusing on the average effect of large school size on large school attendees. In fact, we only need the weaker assumption $Y_i(0) \perp\!\!\!\perp T_i|X_i$ for bias removal. We report subsample sizes for the second survey written test scores (for which the sample size is the largest) according to treatment status and propensity score block in table B.21.

At grade 2, the findings are kept unchanged. Indeed, no effect of large school size has been detected on the test scores, even for the expected more discriminating oral test and the summary index, which has maximum power to detect an effect. This finding holds in the short run and after 18 months, irrespective of the set of covariates used.

The general pattern of the findings at grade 4 is that the use of a particular set of covariates does make the results vary substantially. With the first set of covariates, the average effect of large school size is negative and significant only for the oral test and the index after 18 months. The short-term outcomes are not affected by large school size. The addition of more controls yields estimates that are (again) greater in absolute value and more significant. The effects are detected on all longer-term cognitive outcomes and even on two short-term outcomes (the mathematics outcome and the index). The average effect of large school size on large school attendees is negative and significant for all learning outcomes.

The results from this subsection confirm the long-term and negative effects of large school size on learning outcomes for students of grade 4, as could be anticipated from the previous subsection. By moving through the cycle of elementary school, students' sense of right and wrong evolves accordingly, and they use this knowledge to react to their world; furthermore, awareness of the environment is likely to be greater later in life, probably because the challenges they confront are different at each moment. The ways accumulated knowledge is used and the changes in one's awareness of one's environment at different ages could be reasons why effects are not present at grade 2 but are at grade 4. The absence of large school size effects on cognitive outcomes at the start of elementary school could come from the fact that second graders are taking their first steps in learning and that, possibly, either the instruments are not discriminating enough between students of schools with large or small sizes or the effects of large school size should be observed in non-cognitive outcomes. In the context of Senegal, with about 70 percent of the schools being located in rural areas, most parents are not literate, and even if they are, they often do not write and read the teaching language. Schools tend to be the sole providers of learning, while other relatives, if they can read and write the teaching language, can help with homework. When students start developing their skills, it is not surprising to spot no difference in learning between students in large schools and those in small ones.

Table 2.1: *Estimated ATE and ATET with the Doubly Robust Method by Grade*

		Grade 2				Grade 4			
		Set 1 of Covariates		Set 2 of Covariates		Set 1 of Covariates		Set 2 of Covariates	
		ATE	ATET	ATE	ATET	ATE	ATET	ATE	ATET
French 2010	Estimated effect	-0.020	-0.015	0.013	-0.038	-0.030	-0.085	-0.068	-0.107*
		(0.076)	(0.081)	(0.093)	(0.093)	(0.071)	(0.062)	(0.071)	(0.065)
	Observations	1,409	1,409	407	424	1,321	1,333	434	440
Mathematics 2010	Estimated effect	-0.047	-0.097	-0.048	-0.067	-0.045	-0.066	-0.143*	-0.126*
		(0.060)	(0.061)	(0.075)	(0.071)	(0.070)	(0.060)	(0.075)	(0.073)
	Observations	1,409	1,409	407	424	1,321	1,333	434	440
Oral 2010	Estimated effect	-0.021	-0.003	0.014	-0.014	-0.116	-0.121	-0.137	-0.177**
		(0.075)	(0.081)	(0.093)	(0.095)	(0.080)	(0.081)	(0.092)	(0.088)
	Observations	710	710	351	364	662	667	385	390
Index 2010	Estimated effect	-0.057	-0.056	0.001	-0.042	-0.083	-0.099	-0.125*	-0.149**
		(0.071)	(0.077)	(0.087)	(0.087)	(0.073)	(0.068)	(0.075)	(0.071)
	Observations	702	702	344	357	659	664	382	387
French 2011	Estimated effect	-0.059	-0.115	-0.038	-0.042	-0.091	-0.225***	-0.192**	-0.218*
		(0.090)	(0.094)	(0.116)	(0.114)	(0.087)	(0.085)	(0.094)	(0.111)
	Observations	874	874	267	280	803	814	273	278
Mathematics 2011	Estimated effect	-0.009	-0.071	-0.054	-0.058	-0.076	-0.157*	-0.213**	-0.235**
		(0.081)	(0.081)	(0.099)	(0.097)	(0.087)	(0.090)	(0.096)	(0.109)
	Observations	874	874	267	280	803	814	273	278
Oral 2011	Estimated effect	0.051	0.118	0.155	0.085	-0.291**	-0.265**	-0.339***	-0.383***
		(0.092)	(0.097)	(0.110)	(0.116)	(0.116)	(0.109)	(0.124)	(0.134)
	Observations	435	435	213	221	418	423	236	240
Index 2011	Estimated effect	-0.002	0.029	-0.022	0.004	-0.250**	-0.271***	-0.278**	-0.328***
		(0.097)	(0.103)	(0.119)	(0.113)	(0.107)	(0.096)	(0.109)	(0.110)
	Observations	423	423	207	215	395	400	222	226

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

Robust standard errors accounting for clustering are in parentheses.

The findings, presented in the above table 2.1, also establish that the impacts of large schools mature in the long run, becoming systematically greater and more pronounced over time. Fourth graders that attend large schools perform worse than students in small schools who are otherwise comparable. A large school size hurts even a randomly-selected student after 18 months. These conclusions contradict the theories developed by the proponents of school consolidation. A theory by Ornstein (1990) can help us understand the pattern of the findings. The author stresses that family is the first educational body and the main source of students' development in the early years of schooling. Because young learners, especially those in grade 4, are making the transition from their habitual settings to school in this part of life, a small environment is preferable for their intellectual fulfillment to the seclusion that is associated with large schools.

The results from the doubly robust strategy are consistent with the most recent conclusions on the topic, providing an extension of initial findings to a less developed area. Indeed, Egalite and Kisida (2013) show for primary and secondary schools and Kuziemko (2006) demonstrates for primary schools that school size has an adverse effect on students' outcomes. The effects identified are, however, smaller, probably because more endowments (e.g. more materials, better trained teachers, etc.) are likely to lessen the undesirable effects of school size. Conceptually, small schools are in a better position to offer adequate responses to students' individual needs, as has already been outlined in previous research (Strang, 1987; Maxner, 2005). It is also conceivable that in reasonably-sized schools, where teacher-pupil ratios and attention per pupil are expected to be higher, disruption is less likely to occur. Such conditions seem indispensable for a learning entity to be effectively run and adequately managed. The data show that a higher proportion of principals in large schools consider keeping order to be one of the objectives of the learning community. More specifically, the proportion of school principals that establish discipline as a priority is 11 percentage points lower in the group of small schools. This difference is statistically significant at the stringent level of one percent. Thus, disruption appears to be one of the multiple ways in which school size can hurt learning. Amongst other objectives reported by principals to be more present in the management of small schools are the mastery of the core topics (French and mathematics) and the progression of students through the primary cycle.[23]

[23] We build on the quality of the covariates balance to conduct a plain regression of the test scores on the treatment indicator and the intercept without any conditioning variable and with the weights as constructed for each type of estimator. We use 0.10 for the trimming level. The results from the

2.6.4 Heterogeneity of large school size effects

2.6.4.1 Subpopulations analysis

To gain a better understanding of the extent to which large school size impacts learning outcomes, we explore its heterogeneous effects with respect to some particular covariates (students' gender, the urban/rural location of schools and households' social class). In this subsection, we allow for the effects to be heterogeneous in the population. Formally, we re-estimate equation 2.2 for each subpopulation w, with the constraint that $\Delta_{w,t} = \Delta_t + \mu_{w,t}$ where $\mathbb{V}ar(\mu_{w,t}) > 0$ and $t \in \{1; 2\}$. This heterogeneity analysis is conducted considering only the doubly robust regression (inverse propensity score weighting combined with regression analysis) approach for its desirable properties. To achieve this, we re-estimate the conditional probability of attending a large school for each subgroup in the first stage. As a result of this first stage estimation for subgroups, the weights are different from the ones used when estimating the overall impact.

At grade 2, we note a scarcity of large school size effect that we also observed at the whole school level. The analysis by gender reveals that male students are negatively affected. All effects (except for one positive effect) are insignificant for female students. Rural students are sensitive to large school size in a few cases. For these students, long-term effects are detected only for the mathematics outcome (ATE and ATET). No effect is detected for urban students. Appendix table B.23 presents these findings.

At grade 4, the pattern of the effects is different from what is visible at grade 2: boys and girls tend to be affected after 18 months, even if negative effects are detected in the short run. Large school size does not hurt rural students, whereas urban pupils are the ones who suffer from attending large schools. The estimated impacts in the urban zones are massive and can reach the magnitude of a standard deviation for the oral test after 18 months of exposure. The effects are salient in urban zones, probably because this is where school sizes reflect the most demand for education. These results can be seen in the appendix table B.24.

unadjusted regression agree with those of the doubly robust model for grade 2 in terms of significance of the effects while differences related to the significance of the effects are noted for grade 4. These differences are attributed to the loss of precision that accompanies the non-use of conditioning variables in the plain weighted regression, where the standard errors of the effects are systematically larger than those of the doubly robust regression. See table B.22. Additionally, while the covariates are balanced in the pseudo population, we note differences in the point estimates. This is expected as the doubly robust estimator equals the inverse propensity score weighted estimator plus a second term.

We extend the subgroup analysis to the examination of large school size impacts by social class of households. This analysis is motivated by the fact that prior studies (e.g. Cotton, 1996; Howley, 1996; Caldas, 1993; Franklin and Crone, 1992) suggest that the socioeconomic status of students' households is a mediating factor for the relation between school size and learning achievements. For instance, Franklin and Crone (1992) argued that large school size helps affluent students whereas small schools benefit economically disadvantaged students. A study by Caldas (1993) suggests that learning achievement was unrelated to school size when all schools in Louisiana were considered. However, the results obtained from analyses on a sample restricted to schools in predominantly low socioeconomic areas suggest that larger size was linked to lower achievement. In the same spirit, Cotton (1996) proves that the benefits of small school size were greatest for students from the lower social classes.

To test the hypothesis that the effect of large school size is a function of the social class of students' households, we define two categories of households: those from an affluent family background, who have a wealth index[24] above the median of the distribution of this index, and those with low socioeconomic status, who have a wealth index below the median. We present the results for both grades in the appendix table B.25. The conclusions are that, at grade 2, students from families with a low socioeconomic status learn more when they attend a school with a small size. While these effects are not detected for all outcomes, it is noted that students from wealthy environments are not influenced by school size. At grade 4, students from wealthy origins are negatively affected by school size, but much less than students from poor contexts, who exhibited low outcomes when they attend large schools. However, while poor students experience more severe effects, we note that wealthy students experience more frequent effects.

Again, the grade differential in the impacts of school size is seen here. Our interpretation of these results is that when students start school, their abilities are still more the reflections of home inputs than they are the echoes of the schools' environments. In that sense, high socioeconomic status tends to counterbalance the adverse effects of large school size, whereas low socioeconomic status cannot. This is probably a rationale for seeing negative effects only for impoverished students at this level. At grade 4, however,

[24]The wealth index is defined as the first principal component of the following variables: the home has electricity, the home has plumbing, the home has a radio, the home has a television, the home has a telephone, the home has a computer, the home has a refrigerator, the home has gas, the home has an iron, the home has a bicycle, the home has an automobile, the home has a bed, the home has a modern toilet, etc.

students are, to some extent, homogeneous in terms of abilities as weaker learners tend to drop out; the primary school completion rate was only about 60 percent in Senegal, between 2008 and 2010, according to the Ministry of Education. As students are spending time in schools, family background tends to be less predominant and leaves room for school components, including school size, to affect learning outcomes, regardless of family social class. This result is not new in the literature: the Texas Education Agency (1999) study found a negative effect of increased school size on achievement unrelated to social class.

While it has been shown that large school size has a harmful effect on learning outcomes at grade 4 but not as much at grade 2, the findings from this subsection suggest that the effects of large school size result in fact from interactions between school size and other important variables, such as the grade level, the socioeconomic status of households, the school's location and the student's gender. At grade 2, the observed effects of large schools are mostly driven by rural and quite often poor students. At grade 4, the results are due to students located in urban areas. Rural students at grade 2 are familiar with less crowded surroundings (as rural regions are the less populated regions of the country) and feel uncomfortable in the large school setting. At grade 4, female, male and urban students are affected by large school size and need an environment compatible with the development of their overall maturity.

2.6.4.2 Distributional impact of large school size

The attention we pay to the estimation of quantile treatment effects and quantile treatment effects on the treated stems from their fundamental property of portraying the heterogeneous impact of large school size on any point of the distributions of learning outcomes. This is critically important as, if the mean curve (conditional expectation function) is not informative enough, we may end up understating the overall effects of large school size on learning outcomes.

In this subsection, we estimate unconditional effects rather than conditional effects for the reason stated in the methodology section. Instead of looking at the effect at each point of the distribution of the outcomes, we consider only quantiles of five from five through 95 (5, 10, 15, etc.) and implement the inverse probability weighting method as developed in Firpo (2007). The method is not doubly robust, but it is efficient and semi-parametric. Only the propensity score is estimated, and any functional form of the relation of the outcomes to the covariates is not assumed. We combine this technique with block bootstrap to accurately estimate the standard errors. We resample schools, instead of students, 1,000 times to mimic the sample design in the computation of the standard errors of the estimates.

As learned from previous analyses, the quantile effects are significant only in a very few cases at grade 2, suggesting the absence of an effect of large school size on the distribution of learning outcomes. At grade 4, the harmful effects are more present in the longer-term, which reinforces the idea that length of exposure to school size matters. The adverse effects appear at almost every point of the distribution of students' achievements and are particularly massive.All these results can be seen in tables B.26 through B.29.

2.7 Sensitivity analyses

This section studies the robustness of the results obtained from the analysis conducted above. Until now, we define a large school as a learning entity with size between 440 and 869 students. The threshold (of 440 students) we use so far is based on the distribution of school size in the data, and the previous conclusions of negative effects of large school size rely heavily of this cut-off point. Moreover, we discard from the covariates list those variables that are suspected to be affected (but not much because they have been collected at the start of the school year) by school size. Also, 25 schools we see as too large have been discarded from the main analysis, and two-thirds of our students are not included because of lack of household data for them. We propose to look at the sensitivity of the findings to the modification of the selection of the cut-off point, to the inclusion of supplementary variables, to the exclusion of a set of variables that could have been affected by school size but that we think were important to control for in the main analysis, to the inclusion of data from the schools initially discarded, and to the addition of students with no household data to the estimation sample. These sensitivity analyses are conducted essentially for average effects (ATE and ATET) and outlined below. Finally, we estimate the ATET on a thick support as suggested in Black and Smith (2004).

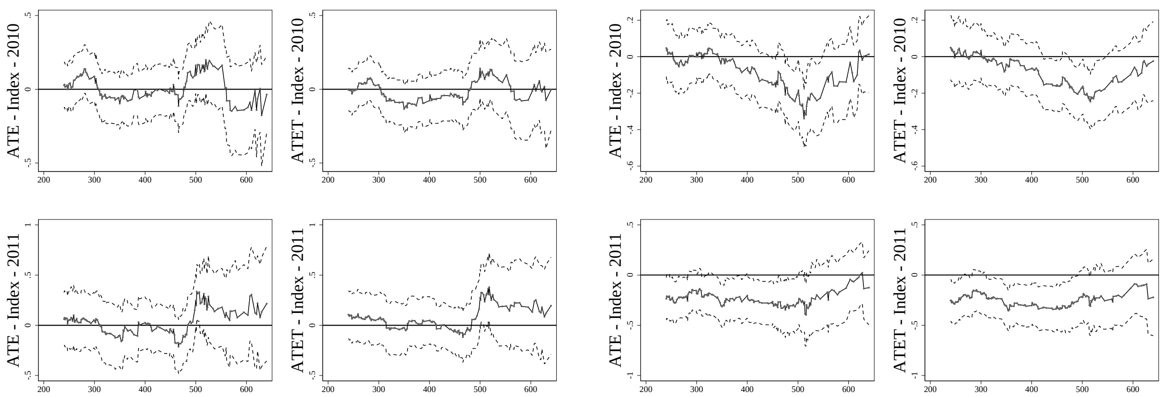
2.7.1 Selection of the cut-off point

In this robustness check, the breakpoint is successively moved from a value to the consecutive one in our data, starting from 240 students and ending at 640. We start at 240 students because this number can be regarded as the size of a standard school of

6 classrooms, each classroom having 40 students on average. A student-teacher ratio of 40:1 is suggested in Mingat et al. (2002) because this ratio is observed in the highest-performing low-income countries. We end at 640 students because, beyond this threshold, the variability of the treatment indicator is seriously deteriorated. Only nine percent of the students are housed in schools with a size greater than 640. Thus, the analysis is operated with a constant bandwidth (200 students) on both sides of our initial cut-off point.

The analyses are done for both estimators (the average effect of large school size on a randomly chosen student and the average effect of large school size on students attending large schools) and for all outcomes (French, mathematics, and oral tests, and the index) at both subsequent surveys. This combination is used as many times as we have incremented the breakpoint point from 240 to 640. Because of jumps in the school size variable, only 147 distinct values of the cut-off point have been studied at grade 2. At grade 4, the data contain 150 distinct values of school size between 240 and 640. We present in figures B.2 through B.4 how the estimates vary as the cut-off point increases when analyzing the French, mathematics and oral tests. We only present here the results for the summary index,[25] which is a weighted average of all other outcomes and expected to reflect the effect of large school size with maximum power.

Figure 2.3: *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes*



A visual analysis of figure 2.3 suggests no impact of large school size on the index for second graders when the cut-off point varies. None of the parameters of interest are

[25]Whenever an average effect is detected at the whole sample level, large school size affects cognitive outcomes in a constant (negative) direction. In such case, a summary index has better power than series of familywise error rate or false discovery rate adjusted individual tests (Anderson, 2008).

statistically significant at grade 2, which is congruent with the absence of effects we have seen with the threshold of 440 students. At grade 4, the estimated effects (ATE and ATET) are significant within a certain range whose bounds determine an interval outside which any effect of any kind is not observable.

A similar sensitivity exercise is performed on the distribution of test scores by means of standard quantile regressions with our covariates as conditioning variables.[26] This robustness check includes not only the original test scores (French, mathematics, and oral), but also their weighted average (the index score). Here, instead of looking at quantiles of five from five through 95 as we did in the main analysis, we look at the entire distribution of test scores at both grades by estimating effects at quantiles of one from one through 99. We then check, for over 235,000 estimates, whether they are positive, negative, or statistically insignificant.[27] The outputs are omitted for space considerations. We note, once more, the paucity of significant effects at grade 2 as 94 percent of the estimates are not statistically significant, 47 percent being related to the short run.[28] Only 2.3 percent of the estimates are positive, and these are mostly for high-achievers (2.2 percent of the estimates relate to high-achievers). Finally, 3.9 percent of the estimates are negative, these adverse effects being mostly detected for high-achieving students (2.4 percent) and almost equally distributed between 2010 and 2011. At grade 4, 81 percent of the estimates are not significant, 45 percent being related to the short run. While only 2.6 percent are positive (2.3 percent for high-achievers), 15.9 percent are negative (10.9 percent for high-achievers and 5 percent for low-achievers). Amongst these harmful effects, 75 percent are detected in the longer-term.

We emphasize here that the proportion of negative effects exceeds the one of positive effects, especially at the fourth grade where it is almost seven times larger. Furthermore, if school size has any positive effects, the positive effects occur in a tiny proportion and mostly relate to top performers, who are, by far, less of a policy concern than weak students. From this analysis, quantile effects exist (especially at grade 4), and the picture

[26] From an econometric standpoint, we estimate the equation $Y_i = \alpha^q + \Delta^q T_i + X_i \delta^q + \varepsilon_i^q$ with $\mathbb{Q}^q(\varepsilon_i) = 0$. α^q, Δ^q and δ^q are the unknown parameters of the model and $\mathbb{Q}^q(\varepsilon_i)$ refers to the *q-th* quantile of the unobserved random variable ε_i.

[27] Since 99 quantiles are analyzed on eight outcomes (French, mathematics, oral and index scores at both 2010 and 2011) and that we consider 147 distinct values of the cut-off point at grade 2 and 150 at grade 4, we check the significance and direction of the 116,424 estimates at grade 2 and 118,800 at grade 4, which represents a total of 235,224.

[28] These percentages do not represent sample fractions. It is the case only if large school size affects all outcomes in the same manner, irrespective of the cut-off selected.

depicted here mimics very well what we saw so far in the main analysis: the importance of length to exposure and the quasi inexistence of effects at grade 2. We refrain from conducting sensitivity analyses using quantile regressions in the following subsections as the similarity of the results we have just discussed to what we have already seen provides a justification for restricting further investigations on mean effects.

The results of the linear and quantile regressions show that the selection of the cut-off point that defines small and large schools does matter and may be part of the reason why the broader literature is inconclusive as to the existence and direction of the effects of large school size.[29]

The fact that the effects vanish beyond a certain threshold is consistent with the deleterious effects of large school size on learning outcomes. In fact, by moving the threshold from left to right, we progressively allow for larger schools to be part of the small schools group. Since large school size harms learning performances (as we demonstrated), the average outcome of the group of schools with small size will decrease as a result of the lower performance of the group of schools with larger size being included in the set of schools that are considered more effective. Furthermore, the evidence that there is no effect below a certain threshold indicates that too small schools, like too large ones, are not effective learning communities. Thus, the interval where large school size negatively affects learning outcomes is a range of school size values that can improve outcomes in the Senegalese education system.

Centered on the summary index and based on the behavior of the estimators of average effects in the short run and after 18 months, we could suggest the "*ideal range*" for school size for the Senegalese education system. This range, which is the intersection of the intervals where an effect has been detected for both estimators (ATE and ATET) for grade 4 on the above figure, is estimated at 470-500 students. At the moment of the data collection, about only 3.3 percent of the students attended a school with a size in the suggested range. The proposed school size fits in any interval based on the index at both grades 2 (second graders are not affected by large school size) and 4. From these last estimates, fourth graders need smaller environment than second graders, and this finding further reinforces the conjecture that the accumulation of knowledge, together with the

[29] Lee and Loeb (2000) define small schools as those with less than 400 students and large schools as those with more than 750 students. According to Conant (1959a, 1967), a large school has a total school population of about 400 students. In the study by Barker and Gump (1964), school size went up to 2,287 students. Lee and Smith (1997) suggest an ideal small high school with about 600-900 students. Lastly, Fox (1981) defined large school as the one having 1,000 students or more.

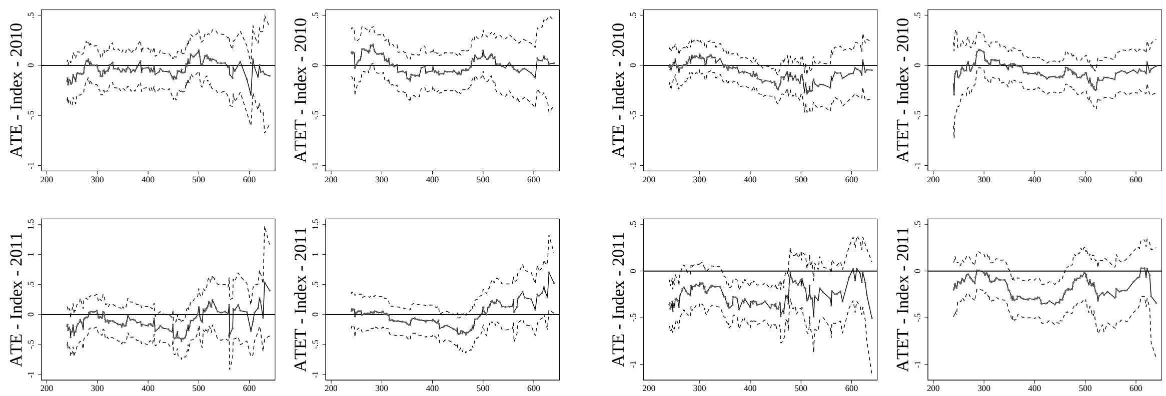
length of exposure, could condition large school size effects.

2.7.2 Addition of size-dependent variables to the set of regressors

The main analysis does not include a set of covariates that has been discarded because they could have been affected by school size. These variables, schools facilities (running water, electrification, library, etc.) and classroom resources and environment (class size, numbers of French and mathematics manuals) may react to school size if principals can anticipate school size. If this is true, then controlling for these variables represents a violation of the assumption of ignorability of treatment (Wooldridge, 2005). However, if the ignorability of treatment still holds, then the inclusion of these variables effectively cancels out any difference due to school resources. The current re-analysis is run for both parameters of interest on the same sample as the main analysis. The effects are looked at on all learning outcomes measured and on the index. We only present the effects on the index. Figures B.5 through B.7 show results for other outcomes. Again, the cut-off point is made to vary from 240 to 640 students.

As can be seen in figure 2.4, the effects are not significant at grade 2 in the short run. They are negative and significant within a minuscule range at grade 2 after 18 months, but they are negative and significant within a wider interval after 18 months at grade 4. For this latter grade, the effects are more intensive in the longer-term, which supports previous findings.

Figure 2.4: *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes*

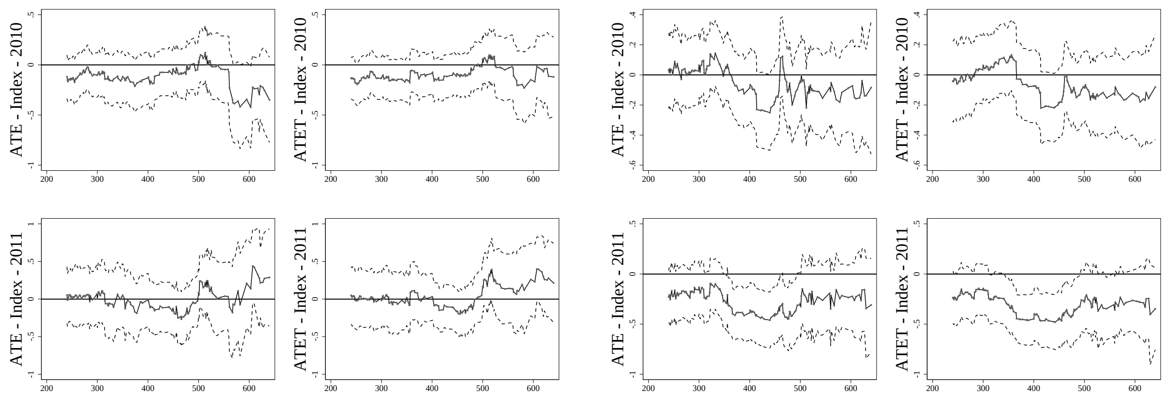


The results from this analysis suggest a range of 350-460 students. Although this range is different from the previous one we found, the lesson is once again that schools of medium size (but not ones that are too large or too small) are a viable policy option.

2.7.3 Exclusion of prior ability from the set of regressors

This analysis is analogous to the previous one, but we run it separately because prior abilities play an important role in our models: they alleviate the omitted variables problem, and, consequently, not including these variables could weaken the assumption of ignorability of treatment. However, prior abilities have been collected concomitantly with other variables we use in this research. Thus, one can argue that these variables are already affected by the treatment, especially if school size reached its actual value some time ago. This is particularly true for fourth graders who have spent more time in school than second graders. While prior ability might pose conceptual problem, it also has a separate, exogenous effect on test scores. We present our results below.

Figure 2.5: *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes*



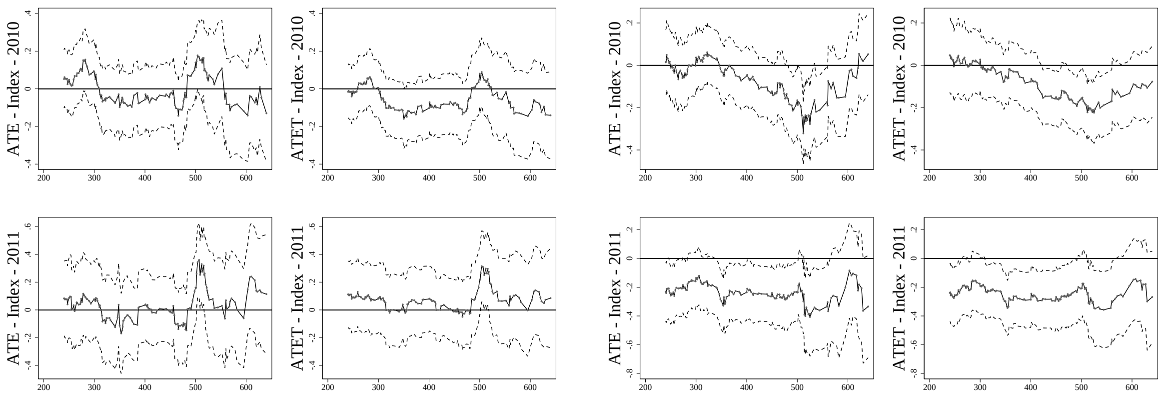
It appears that the effects are not significant at grade 2 in the short run. At grade 4, they are not significant in the short run, but, over a longer period of time, the effects are significant within the range of 350-500 students. Results related to other outcomes are available in figures B.8 through B.10.

2.7.4 Addition of too large schools to the analysis sample

The schools we include here have more than 869 students and represent outliers. However, we exclude the size-dependent variables, with a breakpoint varying from 240 to 640 students. In other words, the only difference between this analysis and the main analysis is the addition of these too large schools.

We run the analysis for both estimators using the index as outcome variable. The results can be viewed in figure 2.6, and the findings from running this sensitivity analysis on other learning outcomes are available in figures B.11 through B.13. It is remarkable that the results are not different from the main analysis findings. At grade 2, there is no effect on learning outcomes, either in the short or the longer runs. At grade 4, the effects are significant within a range of school size, estimated at roughly 340-500 students. As we can see, this sensitivity analysis does not yield contradictory results to what we find in the first robustness check. It is the case because the range suggested here encompasses the 470-500 interval initially obtained.

Figure 2.6: *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes*

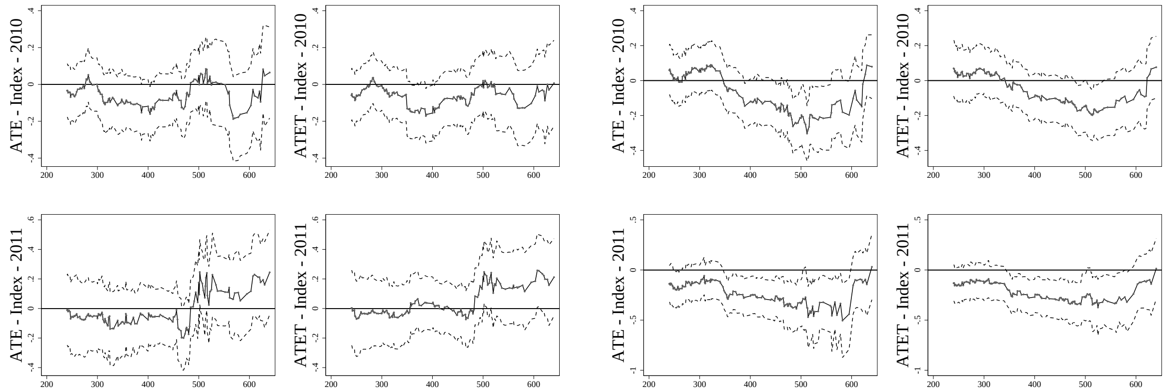


2.7.5 Addition of students with no household data

This analysis is a version of the previous one we ran, but instead of adding schools with too many students, we include students for whom no household data were available due to budgetary constraints. This has the advantage of increasing power, even if slightly. As household interviews were conducted for only a random subsample of students, two-thirds of the original sample is missing at random and thus have missing household characteristics. To be able to keep these observations in the analysis, we assign zero to household characteristics when they are missing and include a dummy indicator for observations with missing household data.

It should be noted that the results of this additional analysis are not different: the longer-term negative effects of large school size are visible for grade 4, whereas no effect can be detected at all for grade 2, as can be seen in figure 2.7. Again, in the longer-term, the interval where significant effects are detected encompasses the interval of 470-500 students. The findings from running this sensitivity analysis on other learning outcomes (French, mathematics, oral) are available in figures B.14 through B.16 in the appendices.

Figure 2.7: *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes*



2.7.6 Minimizing the bias from possible selection on unobservable

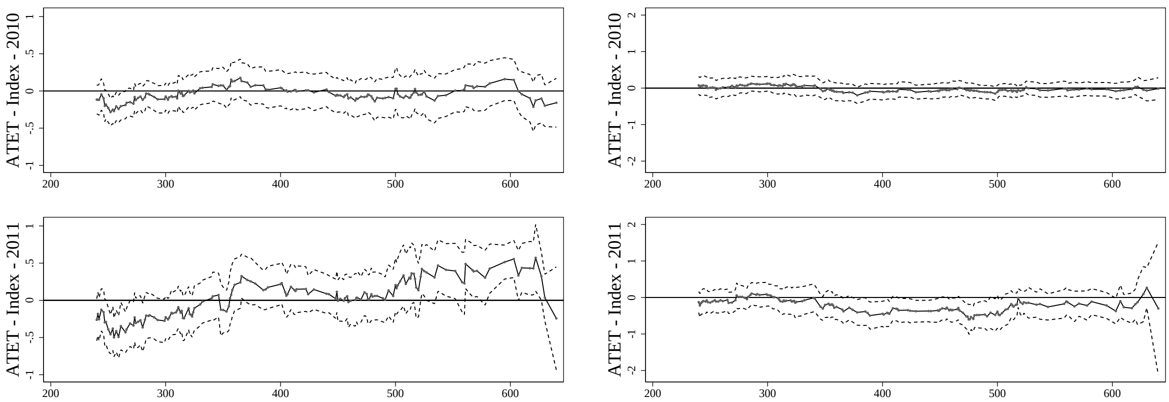
If we think of our results as driven by selection on unobservable variables, we may improve the reliability of the findings by inflating the trimming level (from 0.10 to 0.33) and thus restricting the estimation sample to observations such that the propensity score lies in the interval [0.33-0.67]. As noted in Black and Smith (2004), under the assumptions that (i) potential outcomes and treatment assignment are additively separable in observed and unobserved variables and (ii) there is joint normality of the error term of the comparison group's potential outcome and the error term of the latent variable that determines the treatment, then the bias resulting from a violation of the conditional independence assumption is minimized for the ATET when the propensity score reaches the level 0.50. The authors then propose a sensitivity analysis approach which identifies the ATET that can be estimated with the smallest bias under unconfoundedness.[30] This consists of rerunning the initial identification strategy on observations for which the propensity score lies in between 0.33 and 0.67. While their results rely on joint normality, the intuition behind it is free of this assumption. Indeed, when the probability of selection into treatment is high, unobservable factors on average play a larger role outside the [0.33-0.67] interval than within. In their example, the bias resulting from using the entire treatment group is 7.77 times larger than the bias that remains when the sample is

[30]For estimating the ATE, the bias-minimizing propensity score (BMPS) is not fixed and rather depends on various parameters (see Millimet and Tchernis (2013) for details). Because we are moving the cut-off point from a run to another, the BMPS changes accordingly. Therefore, we are no longer comparing the same population between runs as we restrict the initial sample to observations with propensity scores around this BMPS. For this reason, we refrain from conducting the analysis for the ATE.

restricted to observations in the [0.33-0.67] interval.

We rerun our identification procedure in the [0.33-0.67] interval. Thus, we investigate the consequences of the failure of the conditional independence assumption on the estimated ATET. However, restricting the estimation sample to observations with propensity scores in between 0.33 and 0.67 comes at an expense, specifically that (i) the estimator is less efficient than in previous versions obtained from much larger samples, and (ii) the parameter being estimated is generally different from the population average treatment effect on the treated. The same parameter is estimated only in the case of homogeneous effect, but we see from our heterogeneity analyses that school size interacts with mediating factors, such as the student's gender, school location, etc.

Figure 2.8: *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Index of Measured Cognitive Outcomes*



We present the results from this estimation in the above figure 2.8. Similar results for other outcomes are available in figures B.17 through B.19 in the appendices. From figure 2.8, we can see that the effects are negative only in the long run within small intervals. Amongst these intervals, we identify our initial proposition of 470-500 students per school. Thus, this analysis confirms the robustness of the earlier finding.

2.8 Conclusion

The functioning of primary schools is at the heart of numerous debates on how to improve learning outcomes. These debates are of substantive importance for developing countries that have strongly struggled to expand access to education, in the sense that elementary education is the foundation for subsequent progress in life. Throughout this research that raises the question of school size effects, we have tried to understand the role of school size as one dimension of schooling quality. Such an exercise was challenging not only because of the lack of a precise definition of the acceptable dimensions of a school, but also (and mainly) because empirical estimation in non-randomized settings is subject to strong assumptions. Building on the statement of selection on observable characteristics, we provide evidence of the long-term and detrimental effects of large school size, especially at the higher grades of elementary school. Results from the quantiles models lend some confirmatory evidence of the long-term and adverse effects of school size on learning outcomes. In addition to the grade differential in school size effects that we observe in the main analysis, the study of the heterogeneity of the impacts of large school size suggests that school size interacts with other variables to affect learning outcomes; these variables include the student's gender, the social class of households, and the urban/rural location of the school. Our findings agree with Egalite and Kisida (2013) and Kuziemko (2006). Of major importance, knowledge on the role of school size in the educational process is extended to a developing economy. An emphasis should be put on the fact that the magnitude of the effects is greater in the context of Senegal than those previously examined in the existing literature.

The conclusions of this research support the idea that a school consolidation policy will not necessarily be effective, even if it reduces the operating unit cost and may have other advantages. Harris (2006/2007) argues that the reduction in costs afforded by school consolidation can translate into diminishing returns to scale when the undesirable effects experienced in large schools (less cohesion and the possible need for a more formal bureaucratic structure, which may be costly and inefficient) are larger than those that create increasing returns to scale. Anyhow, attentive management of school dimensions is required to promote students' learning outcomes. Indeed, exceptionally small schools are not advantageous for pupils' performance. As demonstrated above, the addition of students to the community of students in schools whose size was less than a certain number of learners does lead to improved average outcomes. This result makes tenable the assumption that there is an increased probability of enrolling better students by consolidating schools. School consolidation is then affordable up to a certain limit, which our analysis robustly places at 500 students with the lower bound being 470 students given the overall pattern of the sensitivity analyses carried out.

Combining these two findings, the understandings that this investigation provides are entirely congruent with the conclusions of Slate and Jones' review (2005), according to which consolidating schools is initially effective, but the positive gains are inverted as school size keeps rising. Not only is the range of 470-500 students suggested on the basis of partial equilibrium that does not incorporate, for instance, the effects on expenditures, but it should also be noted that the proposed number is founded on the index of the three cognitive outcomes. Summarizing learning outcomes into a single scale has its own flaws, and, as a result, the range of 470-500 students may not be suitable to a particular type of assessment analyzed by itself. Nonetheless, 470 to 500 students could fit in a standard school of 12 classrooms, each classroom having approximately 39-42 students. Given that school retention remains problematic in developing countries and that dropout rates are not yet under control, it would be advisable to enrol more than 42 students in the early grades of elementary school, making it possible to deal with a range of school sizes in a variety of situations. Monk and Haller (1993) rightly pointed out that different schools react to size in different ways, such that it seems impossible to state a single value for the optimal school size that fit all cases. Decision-makers should take into account the characteristics of each community and school before setting the acceptable size.

As a final point, it should be considered that school size, by itself, is only one dimension of school quality. Thus, the goal of establishing policies that aim at improving learning outcomes is achievable not only by dealing with this one aspect of quality, but also by moving on a mixture of strategies that will improve students' performance. Growing economies still face many issues that counterbalance the efforts set up to enhance education quality, and one can improve learning outcomes in many ways. This research has not addressed the cost-benefit analysis of these competitive strategies. It has not taken into account the fact that governments usually have to deal with these different strategies and identify the ones that are the most relevant in such contexts, nor has it studied the effects of school size on other important aspects, such as non-cognitive outcomes. Nonetheless, it has added to the literature, firstly by extending research only available in endowed contexts to a developing country, and secondly by providing a first understanding of an important phenomenon, showing that school size has a substantive and negative impact on learning outcomes after 18 months of exposure to the same treatment, while also highlighting the fact that too small schools tend to have a low level of achievement.

Chapter 3

An Extended Specification of the Three-Level Linear Model

3.1 Introduction

Value-added models have been studied by the *Programme for the Analysis of Education Systems of CONFEMEN* (PASEC) for more than two decades and mostly in Sub-Saharan Africa (SSA). The models account for students' prior achievement values and other background characteristics to measure school and teacher effectiveness as the contribution of school or teacher to students' academic growth. Despite the attractive properties of these models, their relevance for SSA countries has been plagued by a small time lag between the two tests and by disturbances that reduce students' opportunities to learn (strikes, teachers' absenteeism, classes in double shifts, etc.). Because of this small time lag, the effects of classroom-level and school-level variables, as well as those of the extra-school context, are less likely to be detected, resulting in statistical analyses that may suffer from inflated type 2 errors.

Countries participating in PASEC sampled between 150 and 180 schools. Within each school, only one grade 2 and one grade 5 (among those available) classes were selected to be part of the assessment at the two surveys within the same school year. Within the selected classrooms, a maximum of 15 students were randomly sampled. Since only one class per grade was selected, the between-schools variance and the within-schools between-classrooms variance were confounded. Consequently, the value-added models as implemented by the program could not disentangle the effects of context variables at the classroom level from the effects of school-level variables. This type of two-level modelling is common practice in educational research despite the fact that the use of three-level models on achievement data show important clustering effects at the second and third levels of the hierarchy (e.g. Bryk and Raudenbush, 1988; Nye et al., 2004).

It is, however, crucial to integrate the classroom level in the modelling approach when the difference of average achievement between classrooms is significantly different from zero within schools. As widely discussed in the literature (e.g. Van Landeghem et al., 2005; Van Den Noortgate et al., 2005; Moerbeek, 2004), omitting a level in the modelling procedure, like fitting two-level models to three-level data, leads to poor estimation of the variation of the response variable at the two included levels, which in turn affects type 1 errors and prompts inexact conclusions about the relative importance of different sources of influence on the response. Hill and Rowe (1996) suggest that the school-level variance is often reduced to a small proportion of the total variance when classrooms are included as a level between the student and the school; this is due to the fact that the between-classrooms variation in achievement is much larger than that between schools. Another study (Martinez, 2012) indicates that failing to include the classroom level inflates estimates of school-level variance. To overcome these issues and to reflect classroom-level effects independently of school-level effects, we extend a methodology proposed by O'Dwyer (2002) to data collected by PASEC in Francophone Cameroon (Cameroon henceforth), Chad, Congo, Ivory Coast, Senegal, and Togo. Departing from a standard two-level model, O'Dwyer (2002) shows that it is possible to implement a three-level model even if different grades have been surveyed within the same school.

As a reminder, the *Third International Mathematics and Science Study*, known as TIMSS 1995, surveyed 150 schools in participating countries. Within each school, one classroom of grade 7 and one classroom of grade 8 were randomly sampled in most participating countries. O'Dwyer (2002) added the national achievement difference between grade 7 and grade 8 students to all grades 7 classes and considered them as being pseudo-grade 8 classes. As two grade 8 classes are available (a real grade 8 and a pseudo-grade 8) after this procedure, a three-level (student, classroom, and school) regression analysis can be conducted. It was possible in some countries (Australia, Cyprus, Sweden, and the United States of America) to sample two classrooms of the same grade, and these countries have been used as validating countries, where the same methodology has been applied and the results from the real sample were compared to those of the pseudo-samples. It appears from this research that the pseudo-classroom procedure has the potential to be a tenable solution to separating the within-schools between-classrooms variance from the between-schools variance. This methodological approach has been implemented to our data but had to be adapted as grade 2 and grade 5 students were assessed with different instruments and were, therefore, reported on different scales. By considering grade 2 classrooms as pseudo grade 5 classrooms, a three-level model that encompasses all students surveyed in the school can be fitted, instead of having specifications that simply ignore the classroom-level variation in modelling students' learning outcomes.

Existing studies that modelled the same data into a hierarchical linear model perspective include Alognon and Amovin-Assagba (2009), Diagne (2007), Diop (2011) and Michaelowa (2001). Our study extends the specification of the three-level hierarchical linear model and applies it to value-added education production functions. As such, it has a range of advantages over existing research that uses the same data. First, we implement a decomposition of the variance of learning outcomes in three levels (schools, classrooms and students) that better takes into account the structure of educational data; second, under availability of sufficient data at the higher levels (see Maas and Hox, 2005), the proposed specification can easily be extended to have four levels in order to estimate the potential impact of some country-level variables; third, the study of the value-added is undertaken for both a school year and a longer period of three years.

The results of this research are of particular interest. First, the within-schools between-classrooms variation is not negligible and is even highly significant. Second, the three-level specification was successful in isolating important factors of students' academic growth within a single school year and for a longer period of three school years.

The paper is organized as follows. Section 3.2 offers a review of multilevel modelling of learning outcomes, while section 3.3 presents our data. In sections 3.4 and 3.5, we describe the methodological approach undertaken in this paper and discuss our results, respectively. In section 3.6, we conclude.

3.2 A review of multilevel modelling of test scores on PASEC's data

There is an abundant, but still growing, literature relating to the use of multilevel models to understand the patterns and functioning of educational systems. Much of this research looks at the determinants of educational quality (e.g. Michaelowa, 2001), while some of it uses the hierarchical technique to answer specific questions.

The most recent uses of a multilevel model on the PASEC data have been undertaken on four countries: Cambodia, Mali, People's Democratic Republic of Lao (PDR Lao henceforth), and Vietnam (see PASEC, 2014a, b, c, d). The models are constructed as standard growth models and constitute the first attempt by the program itself to use such modelling technique. The use of this approach has led to extensive analyses that include the examination of school effectiveness together with equity, the particular role of some specific variables, and a focus on the variance decomposition, in addition to the search for learning outcomes' determinants in the countries. It was shown that the bulk of the variance of learning outcomes depends on the grade and the country. In Mali, at grade 2 and at grade 5, about 50 percent of the total variance lies at the school level; in PDR Lao, however, the school-level variance is the most important part of the total variance, both at start of the primary cycle and at its end. The profile is fundamentally different in Vietnam, where second graders share an important part of the total variance but the trend is reversed at the end of the elementary cycle, with fifth graders being more homogeneous. The circumstances of the variance decomposition in the case of Cambodia match those described in O'Dwyer (2002). Due to this similarity, the decomposition of the total variance followed three levels: schools, classrooms, and students. Contrary to what Hill and Rowe (1996) found in their research, we find that, in Cambodia, the highest proportion of the total variance is generally attributable to the school-level. The within-schools between-classrooms variance is the least important component in this country.

In response to the importance of socioeconomic status, particular attention was devoted to studying its relation to students' academic achievement. Indeed, as emphasized in Boudon (1973), the economic power of parents generates a range of attitudes, amongst the students, towards school. Bearing in mind the socioeconomic status of their parents, disadvantaged students overestimate the costs of schooling and underestimate its benefits. On the opposite side, students from affluent contexts highly value and appreciate schooling. Wealth seems to generate long-term projects with motivation to complete them, whereas poverty only allows a vision of short-term projects, and even the realization of these seems not to be guaranteed. This supports the idea that students with disadvantaged backgrounds are less successful in school than those from economically wealthy contexts. When investigated, the reports often confirm the positive association between learning outcomes of students and the socioeconomic status of their parents.

However, as has been evidenced in places like Mali, the relation between socioeconomic status and learning outcomes is not actually linear but quadratic instead, suggesting a declining relationship between learning outcomes and socioeconomic status beyond a certain threshold.

Also designed to identify a list of performance factors, the studies reveal that the gender and age of the students, their participation in extra-school activities, teachers' absenteeism, academic background and willingness to serve in the same school, learning conditions in classrooms, classroom equipment, school equipment, etc. are associated with students' attainment. The role teachers' training play is so moderate as to be inexistent (Bernard, Kouak and Vianou, 2005). The results from Diop (2011) are somewhat different. The author re-analyzed the Senegal data into a two-level model perspective and proposed a list of determinants of academic achievement (e.g. the fact that the student possesses a book, the teacher's experience). Other researchers draw conclusions on the basis of a three-level model approach. The third level modelled in the research is, however, common for only two of them. These two studies (Alognon and Amovin-Assagba, 2009; Michaelowa, 2001) consist of cross-countries analysis, which offers the possibility to look at the role of macroeconomic or institutional factors.

Alognon and Amovin-Assagba (2009) study the determinants of learning outcomes for fifth graders in nine countries (Benin, Burkina-Faso, Cameroon, Chad, Congo, Gabon, Madagascar, Mauritania, and Senegal). They used a three-level hierarchical model where students are nested within schools that are located in their respective countries. According to their findings, about 40 percent of the total variance of students' performances is attributable to the first level of the model (i.e. the pupil level) and the same proportion to the school level. Having compared their findings with an alternative ordinary least squares model, they conclude that the results are nearly identical but that the second model tends to overstate the significance of the effects as it does not take into account the hierarchical structure of the data. However, in this case of multiway clustering (students within schools and schools within countries), the authors clustered the standard errors at the country level, thus assuming students are randomly selected within countries, an approach which casts doubts on the findings of the alternative model. Their model also suffers from misestimating the country-level variance (which rests on only nine countries) and not modelling the classroom-level variance. Based on the hierarchical model they proposed, the determinants of learning outcomes include the socioeconomic status of the students' households, participation in extra-school work, repetition, school equipment, teacher experience, the literacy of students' mothers, etc.

Not only Michaelowa (2001) does provide an educational production function, but she also focuses on the efficient use of education-related expenses in five countries (Burkina-Faso, Cameroon, Ivory Coast, Madagascar and Senegal). She concludes that financial resources are much more efficiently used in some countries (Cameroon) than in others. The list of determinants produced at the student level includes the regularity of meals, the students' family background (literacy of parents and the fact that the family uses the French language at home), and the availability of books, radio, and/or television at home. At the school or classroom level, the predictors of students' performances include: 1) teachers' knowledge of the local language, which is positively linked to students' outcomes, 2) class size, which has a U-shaped relation with learning outcomes, and 3) teachers' initial education and regular training, which both play a crucial role. As expected, the expenditures per student play a positive role at the national level, while the illiteracy rate of the country is negatively correlated with test scores. Since the author intended to assess learning achievement during primary education as a whole, and because learning is a cumulative process, the specification did not include the initial value of achievement at the start of the school year. Choosing not to include initial achievement could lead to less clear effects of certain variables, as the author acknowledged.

Clearly, these two efforts in three-level models suffer from a methodological issue that is not discussed in the papers. The number of third level units lies between five (for Michaelowa, 2001) and nine (for Alognon and Amovin-Assagba, 2009). Indeed, one cannot argue that the level-3 variance is correctly estimated with such small numbers of countries.

Amongst multilevel practitioners that use PASEC's data, only Diagne (2007) proposes a three-level model with a sufficient number (61 schools to be exact) of units at the highest level. PASEC's data have been supplemented by newer information on students' households and communities. The specification of the model allows the calculation of the within-students variance (students have been followed-up with annually from 1996 to 2000), the within-schools between-students variance and the between-schools variance. The within-students variance (estimated at 47.5 percent in French and 43.5 in mathematics) was the most important component, whereas the between-schools variance (20.7 percent in French and 16.5 percent in mathematics) accounts for the smallest portion.

The author also proposed a list of determinants of learning outcomes. The determinants suggested in the paper include the parents' educational background, the household's socioeconomic status, the student's age, the student's grade repetition, the teacher's academic background, the class organization mode (whether double shift or multi-grade), etc.

From the description above, the existing works using assessment data collected in SSA clearly show no rigorous attempt to disentangle the school effects from the classroom effects for the purpose of identifying the determinants of learning outcomes. The influence of schools and that of classrooms were pooled together with no possibility of distinguishing between the intrinsic role of class and the real effect of school. A synthesis of PASEC's assessments (see PASEC, 2010) reports schools/classrooms effects as the residuals from the pupil-level ordinary least squares model of learning performances on the space generated by students', teachers', and head teachers' characteristics. Instead, Aitkin and Longford (1986) advocate the use of multilevel models fitted by maximum likelihood method. Likewise, Goldstein (1987) demonstrates the success of multilevel models in estimating the effects of schools policies or practices. Finally, the estimated school effect is partway between two kinds of schools effects[1] recognized in the work of Alwin (1976).

3.3 Data and construction of variables

This study uses data collected by PASEC over the period 2004-2010 in Cameroon, Chad, Congo, Ivory Coast, Senegal, and Togo. Even though PASEC planned to sample up to 180 schools, many of them were finally not surveyed due to circumstances on the ground (e.g. weather conditions, schools' refusals to participate, schools closures, etc.). The program surveyed students at both grades 2 and 5, but, within a small number of schools, one of the grades was not available. Since the methodological approach requires data from a minimum of two grades within each school, we discarded schools with data for one grade only. We also discard the probability of inclusion of students in the sample that is no longer valid since the data organization has changed. For all analyses, second and fifth graders are pooled and compared on a relative scale using the grade equivalency procedure that we will present later in this paper. The sample size per country is given in the following table.

[1]These are named type A and type B effects. See Raudenbush and Willms (1995) for details.

Table 3.1: *Sample Size per Country*

	School year	Number of schools	Number of grade 2 students	Number of grade 5 students	Total number of students
Cameroon	2004-2005	118	1,675	1,683	3,358
Chad	2009-2010	124	1,441	1,525	2,966
Congo	2006-2007	136	1,624	1,848	3,472
Ivory Coast	2008-2009	140	1,891	1,934	3,825
Senegal	2006-2007	134	1,770	1,805	3,575
Togo	2009-2010	170	2,253	2,287	4,540

The data collection, which took place at different time points, includes information at the school or principal levels (age and gender of the principal, academic background and experience of the principal, availability of a library, school resources, working conditions, etc.), teacher- or classroom-level variables (age and gender of the teacher, academic background and experience of the teacher, number of textbooks per student, classroom resources, etc.), and student-level variables (age, gender, language use at home, mother literacy, assignment to extra school activities, family resources, etc.). We have constructed many indices using multiple correspondence analysis. These are the socioeconomic status index, the index of classroom equipment, and the index of school equipment. The socioeconomic status index is built using a set of dichotomous variables that include the material of the walls and roof of the student's house, the assets owned by his/her household (car, television, radio, refrigerator, stove, the type of toilet, and the availability of running water), etc. The classroom equipment index includes the material of the classroom's walls, the availability of a table and chair for the teachers, and other instruction material (such as a blackboard, compass, square, chalk, etc.). The components of the school equipment index are the availability of a functional and equipped library, running water, electricity, an office for the principal, etc. It was possible to aggregate all these dummy variables into a single index as the alpha of Cronbach has a minimum value of 0.60 in all cases. For the purpose of the particular questions we wish to address, we construct many other variables that are the average level of socioeconomic conditions and the standard deviation of this variable at either the classroom or school level. Although the surveys made use of the same instruments for all countries, the datasets do not always contain the same list of variables. For example, the student's gender is missing in the Congo's dataset, whereas the variables that record the student's mother's literacy and use of instructional language at home are not found in the data from Togo. As a result of the unavailability of a number of variables, the models will not be strictly comparable. This is also the case because the data are collected in a six-year window, which is relatively large. We present the list of variables involved in this research, along with their basic characteristics, in the appendix table C.1.

3.4 Empirical approach

In this section, we present the grade equivalency procedure and the econometric approach to the objectives of the research. First, we perform an analysis of variance. In the second stage, we specify two forms of three-level models that are applied to our data. The first specification proposes an educational production function which establishes a list of factors of students' achievement within a school-year, and the second is designed to isolate school-level factors that foster or constrain the relative academic progress between grades 2 and 5.

3.4.1 The grade equivalency procedure

PASEC surveys second and fifth graders in randomly selected schools, and two different scales are built for reporting student performance for these two levels of the elementary cycle. The scales are different because grade 2 and grade 5 tests are different in content and lack anchor items, which in turn prevents the computation of absolute academic progress. In such conditions, students' scores need to be transformed so that the interpretation can be done regardless of the grade at which the test has been implemented. A common way to do this in the literature (e.g. Johnson et al., 2012; Rotz et al., 2014) is to perform a relative comparison by means of standardized test scores. Thus, for each grade and each country separately, we standardize the test scores across schools so that the vectors of learning outcomes all have a mean of zero and a standard deviation of one. This makes the grade 2 and grade 5 test scores in some way comparable as they are now both on a relative single scale. Mapping assessment scales to a unique measure like this allows us to interpret the level of achievement in terms of how close to average students tend to fall regardless of the grade, even if the respective difficulties of our grade-based tests are not the same.

The pre and post-transformation test scores are not numerically equivalent but they rank students over the distributions of learning outcomes in exactly the same way. The transformation operated is just a change in metric as the post-transformation mean score in test score k is given by $\mu_1^k = \hat{\mu} + \frac{\hat{\sigma}}{\sigma^k}(\mu_0^k - \mu^k)$ where μ_0^k is the pre-transformed score in test k, μ^k and σ^k are the original mean and standard deviation of test k, $\hat{\mu}$ and $\hat{\sigma}$ are the desired post-transformation mean and standard deviation for all test scores (respectively 0 and 1 in this paper).

3.4.2 Econometric approach

3.4.2.1 Variance decomposition

The variance decomposition of learning outcomes helps to quantify the variability of test scores at each level of the hierarchy. The estimated model takes the following form:

$$Y_{ijk} = \gamma_{000} + u_k + r_{jk} + e_{ijk} \tag{3.1}$$

where $u_k \sim N(0; \sigma_u^2)$; $r_{jk} \sim N(0; \sigma_r^2)$ and $e_{ijk} \sim N(0; \sigma_e^2)$. Y_{ijk} is the performance of student i, in classroom j, in school k; u_k is the school-level residual, r_{jk} is the classroom-level residual and e_{ijk} is the student-level residual. $\sigma_u^2 = \mathbb{V}ar(u_k)$ is the between-schools variance, $\sigma_r^2 = \mathbb{V}ar(r_{jk})$ is the within-schools between-classrooms variance and $\sigma_e^2 = \mathbb{V}ar(e_{ijk})$ is the variance of Y_{ijk} due to the student level. In such three-level models, two indicators, known as second and third levels intraclass correlation coefficients, are used to describe the variance structure. Also termed variance partition coefficients, these intraclass correlations are respectively the ratios of $\sigma_r^2 = \mathbb{V}ar(r_{jk})$ and $\sigma_u^2 = \mathbb{V}ar(u_k)$ to the total variance $(\sigma_u^2 + \sigma_r^2 + \sigma_e^2)$ of the dependent variable.

The extent to which σ_r^2 is significant indicates how average outcomes of classrooms in the same school are different. This significance provides a justification for a three-level approach rather than a two-level strategy. Once we know the share of the learning outcomes variance between the three levels (student, classroom and school), we perform a second type of decomposition of variance. This decomposition is a conditional analysis of variance where we include, separately, students' initial performance on the one hand and the socioeconomic status of their families on the other. Indeed, we calculate the same proportions as in the first analysis but for students with identical socioeconomic background and similar initial achievement. The reason why we care about conditional analyses of variance is that these two variables play an important role in the learning process. The socioeconomic status of students' parents is an important determinant of later success (Boudon, 1973), and the pre-test is the history of students' schooling before they entered the grade they were attending at the time of the survey.

The general model, assuming all effects at all levels are random, is given as follows:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}X_{ijk} + e_{ijk} \tag{3.2}$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}G_{jk} + r_{0jk} \tag{3.3}$$

$$\pi_{1jk} = \beta_{10k} + r_{10k} \tag{3.4}$$

$$\beta_{00k} = \gamma_{000} + u_{00k} \tag{3.5}$$

$$\beta_{10k} = \gamma_{100} + u_{10k} \tag{3.6}$$

where G_{jk} is the initial grade indicator of classroom j in school k, u_{00k} and u_{10k} are error terms with a multivariate normal distribution with a mean vector of 0 and a covariance matrix T_β. The residual variance $\sigma_e^2 = \mathbb{V}ar(e_{ijk})$ derived from the student-level model described in the above equation 3.2 is computed after controlling for either the socioeconomic status of the student's family or his/her initial ability. In conditional models, the variance partition coefficients are based on the residuals rather than the observed responses. Hence, they measure the proportions of outcome variation that lies at each level of the model and that remains unexplained by the included conditioning variables. However, because equations 3.4-3.6 do not include any predictors, $\mathbb{V}ar(r_{10k})$, $\mathbb{V}ar(u_{00k})$ and $\mathbb{V}ar(u_{10k})$ provide estimations of the variability for the unconditional parameters π_{1jk}, β_{00k} and β_{10k}, respectively.

In theory, the coefficient π_{1jk} of the independent variable varies across schools and can be expressed as a function of the overall slope γ_{100} and a random component u_{10k}. In practice, however, it is essential to assess whether the randomness of the coefficient is supported by the data.[2] To the extent that the coefficient truly varies across schools, it is possible to correlate it with the school mean achievement. In such case, one can assess whether the effect of X is more or less important in high-achieving schools. This type of examination is known as an equity analysis. Indeed, a negative correlation between the school mean achievement and the coefficient of a variable of interest indicates that the effect of X in high-achieving schools is less important than its effect in low-achieving schools, which would suggest that top performing schools are more equitable than those in the lower end of the distributions of learning outcomes. However, this type of equity analysis, although interesting, is beyond the scope of this paper.

[2] The decision to leave the coefficient random could rest on the reliability coefficient computed according to equation 3.58, p. 49 in Raudenbush and Bryk (2002). Reliability estimates smaller than 0.10 indicate that a coefficient considered random should be fixed.

3.4.2.2 A model for the identification of the determinants of educational attainment

Assuming we have student i nested in classroom j that is also nested in school k, we estimate the following fixed effects model:

$$Y_{ijk} = \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk} Z_{ijk} + e_{ijk} \tag{3.7}$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} G_{jk} + \sum_{q=2}^{Q_0} \beta_{0qk} X_{qjk} + r_{0jk} \tag{3.8}$$

$$\beta_{00k} = \gamma_{000} + \sum_{s=1}^{K} \gamma_{00s} W_{sk} + u_{00k} \tag{3.9}$$

The econometric analysis is developed in the framework of a three-level hierarchical linear model. For two arguments, this method is nicer than ordinary least squares. First, hierarchical linear models allow links between different levels to be analyzed. Second, ordinary least squares estimation is unsuitable because the underlying assumption of independent observations is violated. Our proposed technique clarifies the data structure and makes it possible to determine the effects of classroom-level variables independently of the effects of school-level variables.

Equation 3.7 is the level-1 specification and P is the number of predictors. The coefficients π_{pjk} are the level-1 regression coefficients with the corresponding variables $Z's$ being the level-1 predictors; the e_{ijk} are the level-1 residuals with the usual assumption that $e_{ijk} \sim N(0; \sigma_e^2)$.

Equation 3.8 is the level-2 specification. The β_{0qk} are the level-2 regression coefficients, G_{jk} is the dichotomous indicator of the initial grade the student attends at the moment of the survey (1 for grade 5 and 0 for grade 2). The X_{qjk} are the level-2 predictors; the r_{0jk} are the level-2 random effects such that $r_{0jk} \sim N(0; \sigma_\pi^2)$. Because we standardize the two original scales, the average performance of students is zero for each group of students separately, and the grade effect is zero by construction. Thus, it is not expected that the coefficient of G_{jk} is significant in the unconditional models. However, the coefficient of G_{jk} could be significant in conditional models.

Equation 3.9 is the level-3 specification. The γ_{00s} are the level-3 coefficients and the variables W_{sk} are the level-3 predictors, whereas the u_{00k} are the level-3 random effects with the assumption that $u_{00k} \sim N(0; \sigma_\beta^2)$.

To the extent that units of a given level (two or three) are balanced i.e. they have the same numbers of observations, the model can be estimated using generalized least squares. However, in our data, groups are not balanced,[3] and we therefore use a maximum likelihood estimation method.

It is worth mentioning that the above system of equations relates to a mixed model where we link learning outcomes to explanatory variables, including the pre-test. While the motivation of this research is primarily the identification of the determinants of students' progress over a period of three years, whether the pre-test is included in the explanatory variables or not has nothing to do with this objective. The inclusion of the pre-test offers the opportunity to look at a value-added model for a period usually less than a school year, thus matching the working frame used so far, while the exclusion of this variable allows us to look at a model of the test scores as a result of the whole process of schooling since students have started school. Nevertheless, equation 3.8 computes a conditional relative added-value over the period of three years.

3.4.2.3 The longer-term determinants of academic growth

The rationale for conducting this analysis is provided in the introduction of this paper. It is further justified by the fact that, in some countries, schools that are the best performers in early grades are not necessarily the top performers towards the end of the elementary cycle. This remark is a reflection of the differentiated academic progress of students from grade 2 to grade 5, depending on the schools. We illustrate the phenomenon we refer to by correlating the school-level average of grade 2 test scores with the school-level average of grade 5 test scores.[4]

Whereas in Congo the correlation of learning outcomes between grades are as high as 0.74 in mathematics, this correlation falls to 0.39 in Cameroon. The picture for the French score is similar. In Congo, Chad, Ivory Coast, Senegal and Togo, the correlation between grades 2 and 5 French scores is about 0.60. Once more, in Cameroon, it reaches its smallest value, 0.46. This reflects the fact that in Congo, Chad, Ivory Coast, Senegal and Togo, the average second-grade achievement is a good predictor for latter success or failure while, in Cameroon, this is not necessarily the case.

[3]The number of students sampled within grades is between 7 and 15 in Cameroon and Togo, 2 and 26 in Chad, 4 and 15 in Congo, and 6 and 15 in Ivory Coast and Senegal. Each school has two classrooms, so the number of students sampled within the schools varies between 22 and 30 for Cameroon, 4 and 40 for Chad, and 15 and 30 for Congo, Ivory Coast, Senegal, and Togo.

[4]We provide a graphical representation of these correlations in figure C.1.

The goal of this analysis is to determine the pedagogical inputs, extra-school variables, or structural components at the school level that contribute to the relative academic growth of students over a period of three years in the primary cycle. To identify the variables that are associated with the relative value-added between grades 2 and 5, the specification purposefully does not account for student-level or classroom-level variables as in standard education production functions, because we are interested in the full or near-full impact of school variables on students' performances, including any indirect effects that might work through student-level or classroom-level variables. Our focus in this analysis is the model of the slope β_{01k} as a function of school-level variables. The fixed effects specification is formulated as follows, with the notations as in equations 3.7-3.9:

$$Y_{ijk} = \pi_{0jk} + e_{ijk} \tag{3.10}$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} G_{jk} + r_{0jk} \tag{3.11}$$

$$\beta_{00k} = \gamma_{000} + \sum_{s=1}^{K} \gamma_{00s} W_{sk} + u_{00k} \tag{3.12}$$

$$\beta_{01k} = \gamma_{010} + \sum_{s=1}^{K} \gamma_{01s} W_{sk} + u_{01k} \tag{3.13}$$

Also note that the specifications of the equations 3.7-3.9 are different from those of their respective level-correspondents in the set of equations 3.10-3.12. Here, we aim to estimate and model an increase (β_{01k}), whereas the previous specification is interested with the modelling of the values (Y_{ijk}) of test scores. The extra sub-specification (equation 3.13) in the global specification above, represents an important distinction from the previous model and is one of the motivations for this work.

3.5 Results and discussions

3.5.1 Variance decomposition

Here, we provide and discuss the results of the decomposition of the variance into three levels. In previous analyses done by researchers and described above in this paper, the between-schools and the within-schools between-classrooms variance were confounded. The particular interest of this study is to be able to disentangle the within-schools between-classrooms variance from the between-schools variation of learning outcomes.

Table 3.2: *Analysis of Variance on French and Mathematics Learning Outcomes*

		Percentage of the total variance at school level	Percentage of the total variance at classroom level	Percentage of the total variance at student level
Cameroon	French	20,14***	19,16***	60,69
	Mathematics	16,03***	21,02***	62,95
Chad	French	43,28***	22,80***	33,92
	Mathematics	34,26***	22,22***	43,52
Congo	French	48,90***	11,63***	39,47
	Mathematics	43,19***	11,23***	45,58
Ivory Coast	French	30,83***	11,40***	57,76
	Mathematics	24,20***	10,51***	65,29
Senegal	French	30,12***	15,66***	54,22
	Mathematics	23,69***	15,52***	60,79
Togo	French	28,32***	14,31***	57,37
	Mathematics	23,57***	12,02***	64,42

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

According to table 3.2, the student-level variance (i.e. the within-classrooms variance) accounts for more than 50 percent of the total variance in four out of six countries. In Chad and Congo, less than 50 percent of the total variance is accounted for by the pupil-level variance. The variance that lies between classrooms but within schools ranges from 10 to about 25 percent, the largest values being observed in Cameroon and Chad. It shows that the relative performance of the pseudo grade 5 (actually grade 2) students and grade 5 students within schools are significantly different. Finally, the school-level variance usually ranges from about 15 percent to 50 percent. Cameroon presents a profile that is fundamentally different from the other profiles as the school-level and classroom-level variation counts in the same proportion. Cameroon and Chad have the highest proportion of the total variance that is explained at the classroom level, the smallest value of this proportion being observed in Ivory Coast.

It is also useful to note that, for each country, we perform likelihood ratio tests comparing equation 3.1 with a two-level model (where the classroom level is omitted from equation 3.1) of the same type (analysis of variance).[5] The likelihood ratio tests unambiguously rejected the second specifications. The χ^2 with one degree of freedom resulting from these likelihood ratio tests are shown in the appendix table C.2. These empirical χ^2 values, which are for comparison with the critical value 3.84, suggest we reject the null hypothesis that the classroom-level variance can be safely omitted from the

[5]The full (three-level) and restricted (two-level) specifications are fitted using the maximum likelihood estimation method. They are thereby comparable using a likelihood ratio test.

models structure in favour of the alternative hypothesis that this level variance should explicitly be modeled. This is very much consistent with the findings of the variance decomposition.

We note that all level-2 variances are highly significant ($p < 0.001$). A discussion has emerged amongst practitioners of multilevel analysis as to the maximum tolerable p-value of the null hypothesis that the variance of a random component is zero. As stated in Nezlek (2008), the use of a larger (than usual) significance level (at least 0.10) is recommended when testing the significance of variance of random components because, in most situations, coefficients are conceptually heterogeneous and, as much as possible, the model should reflect this theoretical justification. Nonetheless, when a random error term is evidently not significant (e.g. the p-value is greater than 0.20), it is suggested that one estimates a fixed effects model which ignores the random component of the coefficient. The argument in favor of ignoring the random component is that it seems useless to utilize the information in the data to fit an econometric model that is not empirically supported. Our view is that, even if the dependent variable does not significantly vary at upper levels, the computation of standard errors of any estimate still requires integrating the full hierarchical structure of the sample.

Besides the unconditional analysis of variance, we also perform a range of conditional analyses of variance in order to gather information on the proportion of variance explained by some particular variables. These variables include the initial level of the students' achievement and the socioeconomic status of the students' parents. The initial ability of students is an important component in the process of accumulation of knowledge. The research literature on value-added modelling typically assumes that the initial ability captures all previous inputs into student achievement (Hanushek, 1979). However, it is likely that the estimated initial ability contains measurement error, which will bias the regressions coefficients if not corrected for. To address this issue, we include two initial test scores in the models: one in the same subject as the outcome of interest and one in another subject,[6] as already implemented in Lipscomb et al. (2010). The analysis of variance, conditioned on students' prior achievements, shows that these variables explain a large amount of the variability of test scores at all levels. The proportions of reduction

[6]Potamites et al. (2009) use an alternative strategy to overcome the measurement error. They estimate a two-stage least squares model where the student's prior test score in the other subject serves as an instrumental variable for the prior same-subject test score. By instrumenting for the initial score in mathematics with the initial score in French discipline, the model incorporates information on students' performance on the tests in both subjects to measure students' initial achievement.

in school-level variance of mathematics, due to the initial achievements, range from 38 percent in Cameroon to 83 percent in Ivory Coast for the same topic. For the second subject, the minimum reduction of the school-level variance is 53 percent in Cameroon, whereas the maximum decrease in the same level variance is observed again in Ivory Coast and estimated at 81 percent. It is possible that the explanation power of this set of prior achievements is so massive because it results from many variables, including the educational history of the students and their families' background information. These variables certainly represent important predictors to account for in our models. Contrarily, the individual socioeconomic status explains a very small amount (less than one percent) of the variance at either the classroom or the student level. This is true in all countries and irrespective of the subjects analyzed. The socioeconomic status is, however, able to explain a much larger proportion of the school-level variance, especially for the French test score in Togo, where the maximum decrease (32 percent) of the school-level variance is observed (see appendix table C.3). We also note increases in variances at either the classroom or the student level. This is caused by a negative correlation between individual-level variables and group-level errors and does not invalidate the models.[7]

3.5.2 The determinants of learning outcomes

For the countries studied, we estimate four models for each of the French and mathematics test score in order to examine the academic growth of students over the course of a single year. These models, presented in tables C.4 through C.9, are estimated via maximum likelihood with robust standard errors to account for possible heteroskedasticity. In order to verify the normality assumption of the residuals, we have plotted the quantiles of the residuals against those of a normal distribution for each of the 48[8] estimations run in this subsection. These quantile-quantile plots (a total of 144[9]), omitted due to space considerations, confirm normality of residuals, with the exception of a few models that exhibit slightly heavier tails than would be expected from a normal distribution. We also present the Akaike information criterion (AIC) below the models tables. This statistic

[7]Variables in the conditional analysis of variance have been rescaled. They all have a mean of zero and a standard deviation of one. Thus, the observed increase reflects a true pattern and is not the product of a failure to standardize the modelled variables.

[8]For this subsection only, we estimate four models (excluding the analysis of variance) per subject; the analyses include two subjects per country and a total of six countries.

[9]The normality assumptions have been checked at the student level, the classroom level, and the school level for the 48 models estimated.

is a measure of the quality of each model relative to the other models for the same given set of data. In interpreting the results, lower AICs indicate better specifications. From the models,[10] we identify a number of determinants of learning outcomes, which are subsequently classified into three categories: student-level factors, teacher- and classroom-level factors, and principal- and school-level factors. This analysis reveals substantial differences between countries.

3.5.2.1 Student-level factors

The student-level factors we examine include basic student characteristics (such as age and gender), as well as class repetition, mother's literacy status, participation in homework, and family socioeconomic status. The average age of grade 2 students for the entire sample is between seven and eight years, with substantial country-level variation. In Senegal, the average age of second graders is about eight (8.2) years, while the average age in Chad is approximately nine years old. At the grade 5 level, average student age ranges from 11 to 12 years old across the countries analyzed with the exception of Chad, where the average age is approximately 13 years. These average values are as expected as students generally start the primary cycle at six years old. With the exception of Chad, the models exhibit a positive association between student age and performance in mathematics but a negative association between student age and performance in French. In Chad, we find that older students outperform their younger classmates in both French and mathematics on average. The relation of age to learning outcomes is puzzling. A tentative explanation for the negative association between age and French performance is that older age is often associated with a difficult educational trajectory. Indeed, older students have higher rates of class repetition and are more likely to have participated in extra-school activities (e.g. housework, field work or commercial activities). Previous analyses by PASEC show that students that participate in commercial activities develop skills that foster their mathematical competences. Even if these variables (class repetition and participation in extra-school activities) are already included in the models, it is possible that they also work through students' age, something we do not test in our models.

Our models also establish a negative association between learning outcomes and class repetition. Class repetition has long been a subject of discussion in the literature on the

[10] Unless otherwise specified, we discuss the final models.

economics of education, with many proposing dismissal of the practice. First used as a pedagogical tool to help students with poor performance, the practice has been shown to have an impact on dropout rates as well as economic considerations.

Furthermore, Bernard, Simon and Vianou (2005) highlight the fact that there is no common rule in the repetition practices world-wide. Rather, as argued by Brimmer and Pauli (1971), both dropout and class repetition practices are two phenomena that vary widely based upon educational philosophy, economic conditions, and cultural practices. In African countries, which Bernard, Simon and Vianou (2005) assert have the highest class repetition rates internationally, the practice is used to bridge the gap between low-performing and high-performing students. In Scandinavian countries however, students experiencing learning difficulties continue through the educational cycle with the benefit of specific monitoring (Paul and Troncin, 2004). The extent of class repetition varies even further at the country level. Indeed, 65 percent of students in the Cameroon sample have repeated at least one grade in their academic career. We find a similar rate in Chad (60 percent) and Togo (59 percent), while Congo (48 percent), Senegal (43 percent) and Ivory Coast (42 percent) exhibit lower levels of class repetition. Nonetheless, SSA nations as a whole have very high repetition rates as compared to the rest of the world. Contrary to theoretical arguments for positive returns on the practice of class repetition, extensive research, including our contribution here, show that repetition is almost always negatively associated with learning outcomes, even in developed countries. In Paul and Troncin's (2004) analysis of PISA 2000 data from 32 countries of the OECD, they show that Finland, where students were automatically promoted to the next class level, has the highest achievement levels, while Portugal, which has the highest repetition rate of the sample, was at the lower end of the learning outcomes distribution. A causal model developed by Schwille and Eisemont (1991) to understand the reasons for class repetition posited that student-level characteristics (e.g. age, gender, ethnicity, locality of residence, prior repetition), school characteristics (e.g. management policies, coverage of syllabus) and national policies (e.g. quality of instruction) all exert an influence on class repetition. These affect students' learning, motivation, and self-esteem, in turn having an impact on the levels of enrolment, examination success, dropout rates, and the average time needed for students to graduate.

We also look at the student's gender, a characteristic of particular importance due to its relevance to equity concerns and one which we find to be another determinant of knowledge accumulation. Within our samples, we find slight male dominance, with each country sample containing between 40 and 50 percent female students. In our models, girls perform more poorly on average than boys in mathematics. In French, performance by gender varies by country: in Chad and Togo, boys outperform girls on average, whereas in Cameroon and Ivory Coast, the performance of girls, given the same conditions, is identical to that of boys in French. Furthermore, in Senegal, although the average performance of girls is poorer than that of boys, female students that are taught by female teachers exhibit better outcomes. This observation is not entirely unexpected given past evidence on this topic. For example, Anderson (1988) argued that the gender of the teacher affects the teacher-student reciprocal relation, with male teachers providing less encouragement to female students than female teachers.

The models show that participation in extra-school activities is an additional determinant of learning outcomes. Specifically, we study the relationship between participation in field work or commercial activities and student achievement in French and mathematics. We note here that, while we seek to measure the effects of extra-school activities, which effectively reduce learning time and require a certain physical effort, reported participation in housework may not perfectly match these criteria. Overall, we find that participation in extra-school activities is a common practice for students across countries. Indeed, between 30 and 65 percent of the pupils state they help their parents in field work and between 14 and 31 percent affirm they practice small income-generating activities. The students that help in the field work mostly come from disadvantaged families that live in rural settings. Previously, it has been shown that reliance upon children to perform family-supporting economic activities impedes school participation (Anderson, 1988; Lockheed and Verspoor, 1992). When assessed against our data, we find that the use of students in field work is negatively related to learning outcomes.

Although our models do not exhibit causal relationships, the involvement of students in extra-school activities likely weakens the benefits of schooling. Indeed, families face a trade-off between using their children as a labor force for economically productive activities and sending them to school with expectations of positive returns to education. In this trade-off, the opportunity cost of keeping a child in school is the cash earnings or other income lost from forfeited labor. Surprisingly, even in the cases where schools are nearby and inexpensive, families have to be convinced of the advantages from renouncing children's permanent participation in domestic and economic activities (Lloyd and Blanc, 1996). The alleviation of this opportunity cost is the main idea behind the conditional cash transfer programs in education: compensate families for foregone prospects. It has been shown that families do react to direct declines in the cost of education from reduced user-fees (Barrera et al., 2007) or subsidies (Angrist et al., 2002). Filmer and Schady (2011) argued that a very modest transfer (about two percent of mean household consumption) greatly impacted the school enrolment of girls in the kingdom of Cambodia, and that increasing the size of the transfer (to about 3.5 percent) had no extra effect on enrolment, which points to diminishing marginal returns for the transfer amount. Generally, evaluations of cash transfer policies have found large positive impacts on primary and secondary school enrolment (Fiszbein and Schady, 2009).

The second type of extra-school activity of interest is the involvement of students in commercial activities. While strongly related to the first type of extra-school activity, participation in commercial activities has a more mixed correlation with family social background, which prevents us from drawing general conclusions. In fact, in Cameroon, students that do such extra-school activities mostly originate from poor contexts, whereas, in Chad, Ivory Coast and Senegal, student participation in commercial activities is not correlated with socioeconomic background. Finally, in Congo and Togo, these students often come from wealthy contexts.

All countries studied use French as the instructional language. The practice and use of the instructional language at home certainly offers a better understanding of the material taught at school and could lead to improved outcomes. Our data reveal that across all samples, ten percent of students speak French at home. In Cameroon, Congo, and Ivory Coast, this figure is elevated, with between 20-27 percent of students speaking French at home. When empirically assessed, our models show a significant and positive relation between speaking French at home and French learning outcomes in Ivory Coast. As surprising as it may appear, we find a negative correlation between speaking French at home and mathematics outcomes while controlling for student prior ability. We find no significant relationship in Cameroon, but we do note a positive relation in Congo between speaking French at home and French school performance.

The socioeconomic status of a student's family is positively related to his/her learning outcomes in a few, but not all, countries. In Cameroon, Chad, Ivory Coast (French test score only), Senegal, and Togo, there is no significant association in the final models. While this may be unexpected, this result may be attributable to the fact that the effect of socioeconomic status is already captured by measures of initial ability. Indeed, by running models that exclude student initial ability, we confirm this conjecture. These specifications are not value-added models, but they have the advantage of identifying the determinants of learning outcomes as a result of the whole process of schooling, as student initial performance is not taken into account (Michaelowa, 2001). The models without initial ability also reveal that mother's literacy is positively related to learning outcomes in Chad and Ivory Coast, but negatively related to mathematics performance in Senegal. Behrman et al. (1999) find a causal effect of mothers' schooling on their children's school performance, which operates through home teaching in rural India.

Finally, as one might hope, a student's initial performance level is important, and including such information in the model alleviates the issue of omitting important inputs at the school, classroom, family, or student level, many of which are integrated in the initial achievement levels (Hanushek, 1979). The measures of a student's initial performance are highly significant in both final models.

3.5.2.2 Teacher- and classroom-level factors

Evidence on the effects of teacher characteristics on learning outcomes has produced mixed results. Although many researchers claim that teachers' characteristics have no effect on learning outcomes, others have found that some characteristics do have effects on students' achievement, even if moderate in some cases. Blatchford et al. (2004) claim that teachers' characteristics such as age, experience, level of education, length of employment at their current school, or educational level do not exert a particular influence on any discipline at grades 4 through 6. Kane et al. (2008) suggest modest effects of a teacher's initial certification on student academic attainment. Kane and Staiger (2008) argue that teachers' effects quickly faded, whereas Rivkin et al. (2005) found important effects of teacher quality on student performance. They demonstrate that the effects of increasing teacher quality by one standard deviation produced larger gains than a costly ten-student reduction in class size. When assessed with our data, teacher-level variables rarely appear to be statistically significant in the final model. At the country level, we do find a significant relationship with student's performance for teacher status (in Congo, whether a government employee, a contractual or community teacher or not), teacher's gender (Senegal) and professional qualification and experience (Togo). The paucity of significant variables at the teacher level is concordant with previous analyses suggesting that improving teaching quality does not rest, to a large extent, on teachers' characteristics (e.g. Bernard, Kouak and Vianou, 2005). Similarly, Hanushek (1997, 2003) reports that the vast majority (86 percent) of studies conducted in the United States in 1994 find a statistically insignificant relationship between teacher education level and student achievement. This proportion increases to 91 percent when the studies are restricted to value-added models. Teacher's experience levels show a significant association with student's performance for 34 percent of all of these studies and 44 percent of those with value-added models.

Apart from teacher characteristics, we also investigate classroom-level variables that could in some way affect learning outcomes. We find that class size and the socioeconomic composition of classrooms are important determinants of educational outcome in Cameroon, Congo, Senegal, and Togo.

Although class size is a natural continuous variable, we introduce it in our models as a dichotomous variable by dividing the sample between students attending a class of more than 40 students and students attending a class of fewer than 40 students. The threshold of 40 students was chosen as suggested by Mingat et al. (2002). Although Mingat et al. (2002) recommend that the total number of students in the classroom does not exceed 40, the majority of students in our data attend a class larger than this cut-off point. In our sample, Cameroon had the smallest proportion of students in small sized classes, which is estimated at approximately 23 percent. This proportion was 24 percent in Senegal, 27 percent in Congo, 31 percent in Ivory Coast, 35 percent in Chad, and 40 percent in Togo. Clearly, the majority of students are attending classrooms with a large number of students, which will certainly not be beneficial if one relies on the often-perceived positive gains of small class size. Indeed, we find that, in Cameroon and Togo, students' performance declines when class size exceeds 40. The positive gains of small class size are publicized in the vast literature on the topic (see for example Angrist and Lavy, 1999; Krueger, 1999, 2003; Konstantopoulos, 2008, Shin and Raudenbush, 2011) and supported analyses that exploit a causal design. However, Michaelowa (2001) suggests that there is a minimal quality-quantity trade-off when manipulating class size. Indeed, she estimates coefficients indicating that, in a class of 80 pupils, ten additional students would lead to a reduction in achievement by only 1.25 percentage point. At the sample's maximum class size of 139 students, a further increase by ten students would induce a decrease in learning achievement of less than five percentage points (ten percent of average achievement).

The effects of socioeconomic status of students' families on school performance are also tested at the aggregate level and in two different ways: we first look at the compositional effects of socioeconomic status on student performance, and then we examine how homogeneity of socioeconomic status within a class, as measured by standard deviation, is associated with the dependent variables.[11] Findings indicate that the socioeconomic status of peers is positively related to learning outcomes in Cameroon, Chad, Congo, Ivory Coast, Senegal, and Togo. These results are congruent with the existing literature on the topic (e.g. Hoxby, 2000). Alternatively, diversity in classroom socioeconomic composition is correlated with learning outcomes only in Congo, where the association is strong and negative for the French and mathematics outcomes. No significant linear relation is found in the other countries, and this turns out to be the case for both outcomes.

3.5.2.3 Principal- and school-level factors

The effects of the principal's characteristics and school-level variables are sporadic. Among the significant relations, one can distinguish a negative correlation between student achievement and the principal's experience (Ivory Coast), the principal's status (in Senegal, whether civil servant or not), and school size (Congo and Ivory Coast), while we find a positive relationship between student achievement and school equipment (Ivory Coast and Senegal). No relationship between the principal's academic background and student achievement was detected in the analysis.

As we just said, school size is negatively associated with school performance in Ivory Coast and Congo. This finding from our analysis is supported in the literature by a large number of similar studies. However, in the other countries included in this study, no statistically significant relationship is found between student achievement and school size. This highlights the fact that there is indeed mixed evidence on the association between school size and learning outcomes. However, rigorously-conducted causal analyses (e.g.

[11] The analysis of compositional (or contextual) effects occurs with significant regularity (Raudenbush and Bryk, 2002) in endowed contexts where studies of the effects of ability grouping or social segregation are also widely conducted (e.g. Coleman et al., 1966; Burgess et al., 2007; Jenkins et al., 2006). These effects occur when the aggregate of an individual-level characteristic is related to the outcome of interest. The central idea behind the concept of "*compositional effect*" is that the school mean of the covariate affects the association between the student-level covariate and the performance of that student. Shin and Raudenbush (2010) state that, first, school-level average value of the covariate captures characteristics of peers. Second, peers composition may be linked to the resources available to the school and its environment, and these factors in turn may be associated with the student's academic outcome. In the educational literature, this also refers to "*joint production*" as peers' outcomes participate in the student-level outcome value.

Kuziemko, 2006; Schwartz et al., 2011) tend to suggest undesirable effects of school size.

As common sense would suggest, school resources, as measured by school infrastructures and pedagogical materials, tend to raise student achievement. This positive effect is in line with the extensive review by Glewwe et al. (2013), which asserts that the majority of studies show beneficial effects of school infrastructures and pedagogical materials.

Finally, we find that student achievement is still often impacted by the rural or urban status of the school, a fact which highlights the lack of equity and equality of opportunity in educational systems. However, the particular effect of school location status varies by country. In Congo and Togo, rural students perform more poorly than urban ones, whereas rural students outperform their urban counterparts in Chad and Senegal. In Cameroon and Ivory Coast, we detected no significant difference in performance between urban and rural students. Such variability in the relation of learning outcomes to the location of schools is also found in a comparative study of Woessmann (2004), where, in some cases, students in remote schools were found to perform the best.

3.5.3 Slope as an outcome variable: the determinants of academic progress between grades 2 and 5

To measure the determinants of academic progress between grades 2 and 5, we employ a model of slope, which in this case is the grade dummy coefficient. This model offers the advantage of looking at the determinants of student academic growth over a three-year period, which allows for the detection of the full or near-full effect of variables. One drawback of the long observation period, however, is that effects may vanish, and thus may not appear in the estimated models. In this approach, the variables list includes only principal's characteristics and school-level variables, for the reasons stated in the methodology section. The models that link student performance to the variables of interest are presented in tables C.10 and C.11. We find that, in Cameroon, students progress more quickly in schools led by female principals. The average socioeconomic status of a school, as well as an increase in this variable, is also positively related to student academic growth. In Chad and Congo, rural students tend to progress faster than urban pupils. However, the principal's experience level and the existence of a management committee seem to impede academic growth between grades 2 and 5. We also find the surprising result that, in Congo, the school equipment index is negatively associated with academic progress between grades 2 and 5.

3.6 Conclusion

Education plays a central role in the social and economic development of countries (Vandenbussche et al., 2006; Mankiw et al., 1992; Lau et al., 1991). While SSA countries have made efforts to expand access to public education, educational quality has dropped, and large disparities continue to exist in access to a quality education. This research's contribution lies in extending specifications to three-level models; these specifications have so far only been studied either using ordinary least squares or two-level hierarchical linear models. It also describes the relation between various school or extra-school variables and learning outcomes. Our hope is that this investigation will inspire similar studies on newer data so as to better inform stakeholders in education on the operation of educational systems in SSA.

The effects of omitting a level in hierarchical modelling on variance estimates are well-documented in the literature. It has been shown that such omissions lead to misattribution of the response variations and weaken the inferential framework. Nonetheless, the education production functions estimated in the SSA context have thus far ignored the classroom level because only one level was surveyed per grade. By extending the grade procedure proposed in O'Dwyer (2002), this research has opted to re-analyze assessment data from a perspective that better mimics the sampling design. We find substantial and highly significant variation in French and mathematics performance between students of the pseudo-grades within the same school. Our standard approach, which looks at student academic growth within one school year, has been extended to a model that offers the opportunity to model student progress for a longer period, specifically three years for this study (grades 2 to 5). We find that student academic progress is more rapid in rural settings and in schools led by female principals. This academic progress rate is also positively correlated with the school's average socioeconomic status, as well as a change in this variable. Conversely, we find that both the principal's experience level and the existence of a management committee are associated with slower academic growth between grades 2 and 5.

General Conclusion

Policy makers and education stakeholders seek mechanisms to supply higher quality education and in that way improve the educational performance of students. The scientific literature in the field of economics of education has proposed a number of solutions. However, these solutions depend on the intrinsic nature of the education systems, and there is unfortunately no universal answer. In general, correlational analyses at best contributed to the identification of the widely used and so-called best practices. This dissertation goes beyond the flaws of correlational studies in the first two papers presented; causal research methods are implemented as the main support for an evidence-based decision-making process. The third paper is not developed in the causal inference framework but capitalizes on available assessment data to propose an econometric specification that produces trustable inferences and disentangles the effects of school-level variables from classroom-level ones.

In this dissertation, the effectiveness of the Senegalese school grants program, which ran until recently, has been studied. The school grants program has been gauged using the gold-standard approach in the field of causal inference: a randomized control trial. Although participating schools were required to meet the eligibility criterion as defined by the local educational authorities to obtain the grants, we show that our sample is not radically dissimilar from a nationally representative sample. The program has stopped, admittedly, but our results support the decentralization of school resources as one of the ways to go. An important question is whether the school grants were and were perceived to be a permanent or a temporary component of the policy scene. If schools considered them to be a permanent part of the landscape, then they would have taken this into account in deciding how to invest the grant, with the expectation that they would have benefited from another wave of funding. If, however, schools grants were viewed as a one-off occurrence, then school managers would have tried to invest them in a durable way. This would have had lower, but more permanent, effects. Our view from these results is that a well-targeted program that improves schools' resources is likely to have important effects on pupils' performance if it represents a permanent increase in school spending. A short, one-shot program is unlikely to have any longer term effects.

The second chapter of this dissertation provides insights on the functioning of the Senegalese education system with regard to how small schools operate compared to large schools. This was an open question in the context of Sub-Saharan Africa until this dissertation. While being the second causal study of this dissertation, it faced the main challenge of adequately modelling the propensity score to achieve balance on the observable characteristics, assuming that, given the covariates used, any unobservable variable is not related to the treatment assignment. The assumption of selection on the observables remains untestable in our context, but the doubly-robust method implemented makes the results credible. The findings of this study are consistent with other studies (e.g. Kuziemko, 2006; Bloom et al., 2010; Schwartz et al., 2011) that use either instrumental variable estimation or randomized evaluation. The estimated effects are especially important in Senegal, possibly because more endowments in schools contribute to the alleviation of the adverse effects of school size. Not only has this elucidated the role of school size in the learning process, but also, at least as importantly, it has been possible to set the ideal range for school size given the overall behavior of our estimators. The main analysis and the robustness checks suggest that large school size has harmful effects on learning outcomes towards the end of elementary education. It is nevertheless essential to stress that the effectiveness of schools does not solely rest on one dimension, like school size. School effectiveness depends upon many others factors, some of them being related to the management of education systems. In general, the research on school effectiveness suffers from a lack of rigorous causal analyses due to the practical difficulty of implementing either a randomized evaluation or finding a trustworthy instrumental variable.

Other interesting topics that merit further investigation and could guide the direction of the research in the near future include the analysis of class size and class composition and their relations to learning outcomes. The disentanglement of these two effects is also a challenging, but promising, avenue to explore. Michaelowa (2001) suggests no important quantity-quality trade-off in manipulating class size up to 100 students or so, whereas Mingat et al. (2002) suggest a number far lower than 100 (40 students per classroom to be exact). Recently, the frequently suggested policy of class size reduction has been questioned by Duflo et al. (2011), who assert that it is less class size that matters, but more the classroom composition in terms of the innate ability of students. This adds to the contradictory debate between Hanushek (1999a, b) and Krueger (1999, 2003).

As we reported in the general introduction of this dissertation, the research on school effectiveness does suffer from a discordance between the sampling designs and the econometric nature of the models fitted on the data. We develop and test, with a simple analysis of variance, an econometric specification that effectively mirrors the sampling design employed by PASEC to collect the achievement data. These specifications turn out to be better fits on our data than comparable two-level models. As in O'Dwyer (2002), the findings from this simple test pave the way for drawing models that do not omit a level in hierarchical modelling for practical purposes. The proposed specification has also been used to estimate value-added education production functions, which deliver counterintuitive results in some cases and have been able to study pupils' academic growth over a school year and for a longer period of three years. Two of this third essay's conclusions are that students' academic progress is faster in rural settings, in schools headed by female principals and in those with high average socioeconomic status. However, both the principal's experience level and the existence of a management committee are associated with less rapid academic growth between grades 2 and 5.

As it includes only six countries, this research paper is limited firstly in its scope and secondly in that no comparison is possible between the estimated specifications, a circumstance that will not permit one to rigorously infer conclusions for countries on the basis of what is observed in others. The fact that this study is correlational and does not have any causal interpretation is also one reason for prudently drawing lessons on the relation between variables. However, as a regression-based study, it does have descriptive and predictive powers for each of the countries studied.

Bibliography

[1] Abadie, A. and Imbens, G. 2008. *On the failure of the bootstrap for matching estimators*. Econometrica, Vol. 76, No. 6, pp. 1537-1557.

[2] Aitkin, M. and Longford, N. 1986. *Statistical modelling issues in school effectiveness*. Journal of the Royal Statistical Society, Series A (General), Vol. 149, No. 1, pp. 1-43.

[3] Alognon, A. D. and Amovin-Assagba, E. K. T. 2009. *Analyse multiniveaux des facteurs de performances scolaires des élèves du CM1 : Cas des données agrégées des PASEC VII et VIII*. Retrieved February 2014 from *http://www.varlyproject.com/pdf/Analyse%20multiniveaux%20des%20facteurs%20de%20performances%20scolaires%20des%20e%CC%81le%CC%80ves%20du%20CM1.pdf*.

[4] Altonji, J., Elder, T. and Taber, C. 2005. *Selection on observed and unobserved variables: Assesing the effectiveness of catholic schools*. Journal of Political Economy, Vol. 113, pp. 151-184.

[5] Alwin, D. F. 1976. *Assessing school effects: Some identities*. Sociology of Education, Vol. 49, pp. 294-303.

[6] Anderson, M. 2008. *Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects*. Journal of the American Statistical Association, Vol. 103, No. 484, pp. 1481-1495.

[7] Anderson, M. B. 1988. *Improving access to schooling in third world: A review*. Bridge Research Report Series 1.

[8] Angrist, J. D., Lang, D. and Oreopoulos, P. 2009. *Incentives and services for college achievement: Evidence from a randomized trial.* American Economic Journal: Applied Economics, Vol. 1, No. 1, pp. 136-163.

[9] Angrist, J. D. and Lavy, V. 2009. *The effects of high stakes high school achievement awards: Evidence from a randomized trial.* American Economic Review, Vol. 99, No. 4, pp. 1384-1414.

[10] Angrist, J. D., Bettinger, E., Bloom, E., King, E. and Kremer, M. 2002. *Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment.* American Economic Review, Vol. 92, pp. 1535-1559.

[11] Angrist, J. D. and Lavy, V. 1999. *Using maimonides' rule to estimate the effect of class size on scholastic achievement.* The Quarterly Journal of Economics, Vol. 114, No. 2, pp. 533-575.

[12] Angrist, J. D. 1998. *Estimating the labor market impact of voluntary military service using social security data on military applicants.* Econometrica, Vol. 66, No. 2, pp. 249-288.

[13] Banerjee, A. V., Cole, S., Duflo, E. and Linden, L. 2007. *Remedying education: Evidence from two randomized experiments in India.* The Quarterly Journal of Economics, Vol. 122, No. 3, pp. 1235–1264.

[14] Bardhan, P. and Mookherjee, D. 2005. *Decentralizing antipoverty program delivery in developing countries.* Journal of Public Economics, Vol. 89, No. 4, pp. 675-704.

[15] Barker, R. and Gump, P. 1964. *Big school, small school.* Palo Alto, CA, Stanford: University Press.

[16] Barnett, R., Glass, J. C., Snowden, R. and Stringer, K. 2002. *Size, performance, and effectiveness: Cost-constrained measures of best-practice performance and secondary-school size.* Education Economics, Vol. 10, No. 3, pp. 291–311.

[17] Barrera-Osorio, F., Leigh, L. L. and Urquiola, M. 2007. *The effects of user fee reductions on enrollment: Evidence from a quasi-experiment.* Columbia University Department of Economics Mimeo.

[18] Barro, R. J. and Sala-i-Martin, X. 1995. *Economic growth.* New York: McGraw-Hill.

[19] Beare, H., Caldwell, B. J. and Millikan, R. H. 1989. *Creating an excelling school.* New York: Routledge.

[20] Behrman, J. R., Foster, A. D., Rosenzweig, M. R. and Vashishtha, P. 1999. *Women's schooling, home teaching, and economic growth.* Journal of Political Economy, Vol. 107, No. 4, pp. 682-714.

[21] Bernard, J.-M., Simon, O. and Vianou, K. 2005. *Le redoublement : Mirage de l'école africaine?* Dakar: CONFEMEN.

[22] Bernard, J.-M., Kouak, B. T. and Vianou, K. 2005. *Profils enseignants et qualité de l'éducation primaire en Afrique subsaharienne francophone : Bilan et perspectives de dix années de recherche du PASEC.* Retrieved July 2014 from *http://rds.refer.sn/sites/www.confemen.org/IMG/pdf/papier_profils_enseignants.pdf*.

[23] Birdsall, N. and Londono, J. L. 1997. *Asset inequality matters: An assessment of the World Bank's approach to poverty reduction.* American Economic Review Papers and Proceedings, Vol. 87, No. 2, pp. 32-37.

[24] Black, D. and Smith, J. 2004. *How robust is the evidence on the effects of college quality? Evidence from matching.* Journal of Econometrics, Vol. 121, pp. 99–124.

[25] Blatchford, P., Bassett, P., Brown, P., Martin, C. and Russell, A. 2004. *The effects of class size on attainment and classroom processes in english primary schools (Years 4-6) 2000-2003.* Institute of Education, University of London.

[26] Blimpo, M., Evans, D. and Lahire, N. 2014. *School-based management and educational outcomes: Lessons from a randomized field experiment.* The World Bank, Washington, DC.

[27] Bloom, H. S., Thompson, S. L. and Unterman, R. 2010. *Transforming the high school experience: How New York city's new small schools are boosting student achievement and graduation rates.* New York: MDRC.

[28] Boudon, R. 1973. *L'inégalité des chances. La mobilité sociale dans les sociétés industrielles.* Paris, A. Colin.

[29] Bradley, S. and Taylor, J. 1998. *The effect of school size on exam performance in secondary schools.* Oxford Bulletin of Economics and Statistics, Vol. 60, No. 3, pp. 291–324.

[30] Brimmer, M. A. and Pauli, L. 1971. *Wastage in education: A world problem.* Paris, UNESCO.

[31] Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J. and Sturmer, T. 2006. *Variable selection for propensity score models.* American Journal of Epidemiology, Vol. 163, No. 12, pp. 1149–1156.

[32] Bruns, B., Filmer, D. and Patrinos, H. A. 2011. *Making schools work: New evidence on accountability reforms.* The World Bank, Washington, DC.

[33] Bryk, A. S. and Raudenbush, S. W. 1988. *Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model.* American Journal of Education, Vol. 97, No. 1, pp. 65-108.

[34] Burde, D. and Linden, L. 2013. *Bringing education to Afghan girls: A randomized controlled trial of village-based schools.* American Economic Journal: Applied Economics, Vol. 5, No. 3, pp. 27-40.

[35] Burgess, S., Propper, C. and Wilson, D. 2007. *The impact of school choice in England: Implications from the economic evidence.* Policy Studies, Vol. 28, No. 2, pp. 129-143.

[36] Caldas, S. J. 1993. *Reexamination of input and process factor effects on public school achievement.* Journal of Educational Research, Vol. 86, pp. 206-214.

[37] Caldwell, B. J. 2005. *School-based management.* Education Policy Series. Paris: UNESCO, International Institute for Educational Planning and International Academy of Education.

[38] Callahan, R. E. 1962. *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools.* Chicago: University of Chicago Press.

[39] Cheng, Y. C. 1996. *School effectiveness and school-based management: A mechanism for development.* London: Falmer Press.

[40] Chimombo, J. 2005. *Issues in basic education in developing countries: An exploration of policy options for improved delivery*. Journal of International Cooperation in Education, Vol. 8, No. 1, pp.129-152.

[41] Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F. and York, R. 1966. *Equality of educational opportunity*. Washington, DC: United Stated Government Printing Office.

[42] Conant, J. B. 1967. *The comprehensive high school*. New York: McGraw-Hill.

[43] Conant, J. B. 1959a. *The American high school today*. New York: McGraw-Hill.

[44] Conant, J. B. 1959b. *The child, the parent and the state*. Cambridge, MA: Harvard University Press.

[45] Cotton, K. 1996. *School size, school climate, and student performance*. Portland, OR: Northwest Regional Educational Laboratory.

[46] Crispin, L. 2011. *School size and student achievement: Does one size fit all?* Retrieved February 2013 from *http://web.econ.ohio-state.edu/~crispin/docs/Crispin_SchoolSizeandStudentAchievement_11-1-11.pdf*.

[47] Crump, R. K., Hotz, V. J., Imbens, G. W. and Mitnik, O. A. 2009. *Dealing with limited overlap in estimation of average treatment effects*. Briometrika, Vol. 96, No. 1, pp. 187-199.

[48] DeStefano, J., Lynd, M. R. and Thornton, B. 2009. *La qualité de l'éducation de base au Sénégal : Une revue*. Rapport Final. Prepared for USAID Senegal. Retrieved November 2012 from *http://pdf.usaid.gov/pdf_docs/PNADP649.pdf*.

[49] Diagne, A. 2007. *Déterminants des apprentissages dans l'éducation primaire au Sénégal*. Consortium pour la Recherche Economique et Sociale, Série de documents de recherche, No. 2007/01.

[50] Diop, A. 2011. *Politique de recrutement des enseignants non fonctionnaires et qualité de l'éducation de base au Sénégal : Quels enseignements vers l'Education Pour Tous (EPT)?* Retrieved January 2014 from *http://tel.archives-ouvertes.fr/docs/00/62/07/82/PDF/11022.pdf*.

[51] Doksum, K. 1974. *Empirical probability plots and statistical inference for nonlinear models in the two-sample case.* The Annals of Statistics, Vol. 2, No. 2, pp. 267-277.

[52] Duflo, E., Dupas, P. and Kremer, M. 2011. *Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya.* American Economic Review, Vol. 101, No. 5, pp. 1739-1774.

[53] Dynarski, S. 2008. *Building the stock of college-educated labor.* The Journal of Human Resources, Vol. 43, No. 3, pp. 576-610.

[54] Education For All–Fast Track Initiative. 2010. *Highlight results on the ground.* Retrieved June 2012 from *http://www.globalpartnership.org/media/library/About%20FTI/Fact_Sheet_September-2010.pdf*.

[55] Education International. 2008. *Education For All by 2015: Education International's response to the global monitoring report 2008.* Retrieved June 2012 from *http://download.ei-ie.org/docs/IRISDocuments/EI%20Publications/Other%20Publications/2008-00169-01-E.pdf*.

[56] Egalite, A. J. and Kisida, B. 2013. *The impact of school size on student achievement: Evidence from four states.* EDRE Working Paper No. 2013-03. Retrieved July 2012 from *http://www.uaedreform.org/wp-content/uploads/2013/03/Egalite-Kisida-13_Impact-of-School-Size-on-Student-Achievement_wp.pdf*.

[57] Filmer, D. and Schady, N. 2011. *Getting girls into school: Evidence from a scholarship program in Cambodia.* Journal of Development Economics, Vol. 96, No. 1, pp. 150-157.

[58] Firpo, S. 2007. *Efficient semiparametric estimation of quantile treatment effects.* Econometrica, Vol. 75, pp. 259-276.

[59] Fiszbein, A. and Schady, N. 2009. *Conditional cash transfers: Reducing present and future poverty.* The World Bank, Washington, DC.

[60] Fox, W. F. 1981. *Reviewing economies of size in education.* Journal of Education Finance, Vol. 6, pp. 273-296.

[61] Franklin, B. J. and Crone, L. J. 1992. *School accountability: Predictors and indicators of Louisiana school effectiveness*. Paper presented at the annual meeting of the Mid-South Educational Association, Knoxville, TN.

[62] Frölich, M. and Melly, B. 2013. *Unconditional quantile treatment effects under endogeneity*. Journal of Business and Economic Statistics, Vol. 31, No. 3, pp. 346-357.

[63] Frölich, M. and Melly, B. 2010. *Estimation of quantile treatment effects with Stata*. Stata Journal, Vol. 10, No. 3, pp. 423–457.

[64] Fuller, B. 1987. *What factors raise achievement in the third world*. Review of Educational Research, Vol. 57, pp. 255-292.

[65] Galiani, S. and Perez-Truglia, R. 2013. *School management in developing countries*. In Paul Glewwe (Eds.), Education Policy in Developing Countries, pp. 193-242. Chicago: University of Chicago Press.

[66] Galiani, S., Gertler, P. and Schargrodsky, E. 2008. *School decentralization: Helping the good get better, but leaving the poor behind*. Journal of Public Economics, Vol. 92, No. 10, pp. 2106–2120.

[67] Garibaldi, P., Giavazzi, F., Ichino, A. and Rettore, E. 2012. *College cost and time to complete a degree: Evidence from tuition discontinuities*. The Review of Economics and Statistics, Vol. 94, No. 3, pp. 699-711.

[68] Glazerman, S., Levy, D. M. and Myers, D. 2003. *Non experimental versus experimental estimates of earnings impacts*. The Annals of the American Academy of Political and Social Science, Vol. 589, pp. 63–93.

[69] Glewwe, P., Hanushek, E. A., Humpage, S. and Ravina, R. 2013. *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010*. In Paul Glewwe (Eds), Education Policy in Developing Countries, pp. 13-64. Chicago: University of Chicago Press.

[70] Glewwe, P., Kremer, M. and Moulin, S. 2009. *Many children left behind? Textbooks and test scores in Kenya*. American Economic Journal: Applied Economics, Vol. 1, No. 1, pp. 112–135.

[71] Glewwe, P. and Kremer, M. 2006. *Schools, teachers, and education outcomes in developing countries.* Handbook of the Economics of Education, Vol. 2, pp. 945–1017.

[72] Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E. 2004. *Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya.* Journal of Development Economics, Vol. 74, No. 1, pp. 251–268.

[73] Goldstein, H. 1997. *Methods in school effectiveness research.* School Effectiveness and School Improvement, Vol. 8, No. 4, pp. 369-395.

[74] Goldstein, H. 1987. *Multilevel models in educational and social research.* London: Oxford University Press.

[75] Hanushek, E. A., Link, S. and Woessmann, L. 2013. *Does school autonomy make sense everywhere? Panel estimates from PISA.* Journal of Development Economics, Vol. 104, pp. 212–232.

[76] Hanushek, E. A. and Woessmann, L. 2010. *The economics of international differences in educational achievement.* Handbook of the Economics of Education, Vol. 3, pp. 89-200.

[77] Hanushek, E. A. 2006. *School resources.* In Eric A. Hanushek and Finis Welch (Eds.), Handbook of the Economics of Education, Vol. 2, pp. 865-908.

[78] Hanushek, E. A. and Rivkin, S. G. 2006. *Teacher quality.* Handbook of the Economics of Education, Vol. 2, pp. 1051–1078.

[79] Hanushek, E. A. 2003. *The failure of input-based schooling policies.* The Economic Journal, Vol. 113, No. 485, pp. F64–F98.

[80] Hanushek, E. A. 1999a. *Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects.* Educational Evaluation and Policy Analysis, Vol. 21, No. 2, pp. 143-163.

[81] Hanushek, E. A. 1999b. *The evidence on class size.* In Susan E. Mayer and Paul E. Peterson (Eds.), Earning and Learning: How Schools Matter, Washington, DC: Brookings Institution, pp. 131-168.

[82] Hanushek, E. A. 1997. *Assessing the effects of school resources on student performance: An update*. Educational Evaluation and Policy Analysis, Vol. 19, No. 2, pp. 141–164.

[83] Hanushek, E. A. 1979. *Conceptual and empirical issues in the estimation of educational production functions*. Journal of Human Resources, Vol. 14, No. 3, pp. 351-388.

[84] Hanushek, E. A. 1972. *Education and race: An analysis of the educational production process*. DC Heath and Company, 285 Columbus Avenue, Boston, Mass. 02116.

[85] Harris, D. N. 2006/2007. *Class size and school size: Taking the trade-offs seriously*. Brookings Papers on Education Policy, No. 9, pp. 137-161.

[86] Hastings, J. S., Kane, T. J. and Staiger, D. O. 2006. *Gender and performance: Evidence from school assignment by randomized lottery*. American Economic Review, Vol. 96, No. 2, pp. 232-236.

[87] Heckman, J. J., Stixrud, J. and Urzua, S. 2006. *The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior*. Journal of Labor Economics, Vol. 24, No. 3, pp. 411-482.

[88] Heckman, J. J. and Rubinstein, Y. 2001. *The importance of non-cognitive skills: Lessons from the GED testing program*. American Economic Review, Vol. 91, No. 2, pp. 145-149.

[89] Heckman, J. J., Ichimura, H. and Todd, P. 1998. *Matching as an econometric evaluation estimator*. Review of Economic Studies, Vol. 65, pp. 261–294.

[90] Hill, P. W. and Rowe, K. J. 1996. *Multilevel modelling in school effectiveness research*. School Effectiveness and School Improvement, Vol. 7, pp. 1-34.

[91] Hirano, K. and Imbens, G. 2004. *The propensity score with continuous treatments*. In A. Gelman and X. L. Meng (Eds.), Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives, pp. 73–84. New-York: John Wiley.

[92] Hirano, K., Imbens, G. and Ridder, G. 2003. *Efficient estimation of average treatment effects using the estimated propensity score*. Econometrica, Vol. 71, No. 4, pp. 1161-1189.

[93] Horvitz, D. G. and Thompson, D. J. 1952. *A generalization of sampling without replacement from a finite universe*. Journal of the American Statistical Association, Vol. 47, pp. 663-685.

[94] Howley, C. 1996. *Compounding Disadvantage: The effects of school and district size on student achievement in West Virginia*. Journal of Research in Rural Education, Vol. 12, No. 1, pp. 25-32.

[95] Hoxby, C. 2000. *Peer effects in the classroom: Learning from gender and race variation*. National Bureau of Economic Research Working Paper No. 7867.

[96] Imai, K. and Van Dyk, D. A. 2004. *Causal inference with general treatment regimes: Generalizing the propensity score*. Journal of the American Statistical Association, Vol. 99, No. 467, pp. 854-866.

[97] Imbens, G. W. and Wooldridge, J. M. 2009. *Recent developments in the econometrics of program evaluation*. Journal of Economic Literature, Vol. 47, No. 1, pp. 5–86.

[98] Imbens, G. W. 2000. *The role of the propensity score in estimating dose-response functions*. Biometrika, Vol. 87, No. 3, pp. 706-710.

[99] Imbens, G. W. and Rubin, D. B. 1997. *Bayesian inference for causal effects in randomized experiments with noncompliance*. Annals of Statistics, Vol. 25, No. 1, pp. 305-327.

[100] Jackson, K. C. 2010. *Peer quality or input quality? Evidence from Trinidad and Tobago*. National Bureau of Economic Research Working Paper No. 16598.

[101] Jencks, C. S., Smith, M., Ackland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. and Micholson, S. 1972. *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

[102] Jenkins, S. P., Micklewright, J. and Schnepf, S. V. 2006. *Social segregation in secondary schools: How does England compare with other countries?* Oxford Review of Education, Vol. 34, No. 1, pp. 21-37.

[103] Johnson, M., Lipscomb, S., Gill, B., Booker, K. and Bruch, J. 2012. *Value-added models for the Pittsburgh public schools*. Retrieved January 2014 from *http://files.eric.ed.gov/fulltext/ED531790.pdf*.

[104] Kane, T. J. and Staiger, D. O. 2008. *Estimating teacher impacts on student achievement: An experimental evaluation*. National Bureau of Economic Research Working Paper No. 14607.

[105] Kane, T. J., Rockoff, J. and Staiger, D. 2008. *What does certification tell us about teacher effectiveness? Evidence from New York city*. Economics of Education Review, Vol. 27, pp. 615–631.

[106] Kling, J. R., Liebman, J. B. and Katz, L. F. 2007. *Experimental analysis of neighborhood effects*. Econometrica, Vol. 75, No. 1, pp. 83-119.

[107] Konstantopoulos, S. 2008. *Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR*. Elementary School Journal, Vol. 108, pp. 275-291.

[108] Krueger, A. B. 2003. *Economic considerations and class size*. The Economic Journal, Vol. 113, pp. F34–F63.

[109] Krueger, A. B. 1999. *Experimental estimates of education production functions*. The Quarterly Journal of Economics, Vol. 114, No. 2, pp. 497-532.

[110] Kuziemko, I. 2006. *Using shocks to school enrollment to estimate the effect of school size on student achievement*. Economics of Education Review, Vol. 25, No. 1, pp. 63-75.

[111] Lau, L. J., Jamison, D. T. and Louat, F. F. 1991. *Education and productivity in developing countries: An aggregate production function approach*. Policy Research Working Paper No. 612. The World Bank, Washington, DC.

[112] Lee, V. E. and Loeb, S. 2000. *School size effect in Chicago elementary schools: Effects on teachers' attitudes and students' achievement*. American Educational Research Journal, Vol. 37, No. 1, pp. 3–31.

[113] Lee, V. E. and Smith, J. B. 1997. *High school size: Which works best, and for whom?* Educational Evaluation and Policy Analysis, Vol. 19, No. 3, pp. 205-227.

[114] Lehmann, E. 1974. *Nonparametrics: Statistical methods based on ranks*. San Francisco: Holden-Day.

[115] Leithwood, K. and Menzies, T. 1998. *Forms and effects of school-based management: A review*. Educational Policy, Vol. 12, No. 3, pp. 325-346.

[116] Li, F., Zaslavsky, A. M. and Landrum, M. B. 2013. *Propensity score weighting with multilevel data*. Statistics in Medicine, Vol. 32, No. 19, pp. 3373-3387.

[117] Lingle, J. 2009. *Evaluating the performance of propensity scores to address selection bias in a multilevel context: A Monte Carlo simulation study and application using a national dataset*. Educational Policy Studies Dissertations, Paper 56.

[118] Lipscomb, S., Gill, B., Booker, K. and Johnson, M. 2010. *Estimating teacher and school effectiveness in Pittsburgh: Value-added modeling and results*. Retrieved June 2014 from *http://www.mathematica-mpr.com/~/media/publications/pdfs/education/pittsburgh_vam.pdf*.

[119] Lloyd, C. B. and Blanc, A. K. 1996. *Children schooling in Sub-Saharan Africa*. Population Development Review, Vol. 22, No. 2, pp. 265-298.

[120] Lockheed, M. and Verspoor, A. M. 1992. *Improving primary education in developing countries: A review of policy options*. The World Bank, Washington, DC.

[121] Maas, C. J. M. and Hox, J. J. 2005. *Sufficient sample size for multilevel modeling*. Methodology, Vol. 1, No. 3, pp. 86–92.

[122] Malen, B., Ogawa, R. T. and Kranz, J. 1990. *What do we know about site-based management: A case study of the literature - A call for research*. In William H. Clune and John F. White (Eds.), Choice and Control in American Education: The Practice of Choice, Decentralization and School Restructuring, Vol. 2, pp. 289–342. London: Falmer Press.

[123] Mankiw, N. G., Romer, D. and Weil, D. N. 1992. *A contribution to the empirics of economic growth*. The Quarterly Journal of Economics, Vol. 7, No. 2, pp. 407-437.

[124] Martinez, J. F. 2012. *Consequences of omitting the classroom in multilevel models of schooling: An illustration using the effects of opportunity to learn on reading achievement*. School Effectiveness and School Improvement, Vol. 23, No. 3, pp. 305-326.

[125] Maxner, N. 2005. *School climate and school size: Implications for the role of the school psychologist*. Unpublished master's thesis. Mount Saint Vincent University, Halifax, Nova Scotia, Canada.

[126] McGaw, B., Piper, J., Banks, D. and Evans, B. 1992. *Making schools more effective*. Hawthorn, Victoria: Australian Council for Educational Research.

[127] Michaelowa, K. 2001. *Primary education quality in francophone Sub-Saharan Africa: Determinants of learning achievement and efficiency considerations*. Retrieved April 2014 from *http://www.ipz.uzh.ch/institut/mitarbeitende/staff/michaelowa/publikationen/Papers/15.pdf*.

[128] Millimet, D. L. and Tchernis, R. 2013. *Estimation of treatment effects without an exclusion restriction: With an application to the analysis of the school breakfast program*. Journal of Applied Econometrics, Vol. 28, No. 6, pp. 982-1017.

[129] Mincer, J. 1958. *Investment in human capital and personal income distribution*. Journal of Political Economy, Vol. 66, No. 4, pp. 281-302.

[130] Mincer, J. 1975. *Schooling, experience, and earnings*. Journal of Political Economy, Vol. 83, No. 2, pp. 444-446.

[131] Mingat, A., Rakotomalala, R. and Tan, J.-P. 2002. *Financing Education For All by 2015: Simulations for 33 African countries*. Africa Region Human Development Working Paper. The World Bank, Washington, DC.

[132] Ministry of Economy and Finance of Senegal and World Bank. 2004. *La pauvreté au Sénégal: De la dévaluation de 1994 à 2001-2002*. Preliminary version. Dakar.

[133] Ministry of Education of Senegal. 2011. *Rapport national sur la situation de l'éducation*. Dakar.

[134] Moerbeek, M. 2004. *The consequence of ignoring a level of nesting in multilevel analysis*. Multivariate Behavioral Research, Vol. 39, pp. 129–149.

[135] Monk, D. H. and Haller, E. J. 1993. *Predictors of high school academic course offerings: The role of school size*. American Educational Research Journal, Vol. 30, No. 1, pp. 3-21.

[136] Moretti, E. 2004. *Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data*. Journal of Econometrics, Vol. 121, pp. 175–212.

[137] Murnane, R. J. and Ganimian, A. J. 2014. *Improving educational outcomes in developing countries: Lessons from rigorous evaluations*. National Bureau of Economic Research Working Paper No. 20284.

[138] Nezlek, J. B. 2008. *An introduction to multilevel modeling for social and personality psychology*. Social and Personality Psychology Compass, Vol. 2, pp. 842-860.

[139] Newman, F. 1992. *Student engagement and achievement in American secondary schools*. New York: Teachers College Press.

[140] Nye, B., Konstantopoulos, S. and Hedges, V. E. 2004. *How large are teacher effects?* Educational Evaluation and Policy Analysis, Vol. 26, pp. 237-257.

[141] O'Brien, P. 1984. *Procedures for comparing samples with multiple endpoints*. Biometrics, Vol. 40, No. 4, pp. 1079–1087.

[142] O'Dwyer, L. 2002. *Extending the application of multilevel modeling to data from TIMSS*. In D. F. Robitaille and A. E. Beaton (Eds.), Secondary Analysis of the TIMSS Data, pp. 359-373.

[143] Ornstein, A. 1990. *School size and effectiveness: Policy implications*. The Urban Review, Vol. 22, No. 3, pp. 239-245.

[144] PASEC. 2014a. *Performances scolaires et facteurs de la qualité de l'éducation en République Socialiste du Vietnam*. Rapport d'évaluation diagnostique au Vietnam 2011/2012. Dakar: CONFEMEN.

[145] PASEC. 2014b. *Performances scolaires et facteurs de la qualité de l'éducation en République Démocratique Populaire Lao*. Rapport d'évaluation diagnostique en République Démocratique Populaire Lao 2011/2012. Dakar: CONFEMEN.

[146] PASEC. 2014c. *Performances scolaires et facteurs de la qualité de l'éducation dans l'enseignement primaire publique au Cambodge*. Rapport d'évaluation diagnostique au Cambodge 2011/2012. Dakar: CONFEMEN.

[147] PASEC. 2014d. *Qualité de l'enseignement fondamental au Mali: Quels enseignements ?* Rapport d'évaluation diagnostique au Mali 2011/2012. Dakar: CONFEMEN.

[148] PASEC. 2010. *Synthèse des résultats PASEC VII, VIII et IX*. Retrieved March 2014 from *http://www.confemen.org/wp-content/uploads/2012/03/Synthese_PASEC_VII-VIII-IX_final.pdf*.

[149] PASEC. 2007. *Diagnostic des apprentissages dans un système éducatif en expansion*. Rapport d'évaluation diagnostique au Sénégal 2006/2007. Dakar: CONFEMEN.

[150] Paul, J. J. and Troncin, T. 2004. *Les apports de la recherche sur l'impact du redoublement comme moyen de traiter les difficultés scolaires au cours de la scolarité obligatoire*. Haut conseil de l'évaluation de l'école, No. 14. Paris.

[151] Potamites, L., Booker, K., Chaplin, D. and Isenberg, E. 2009. *Measuring school and teacher effectiveness in the EPIC charter school consortium – Year 2*. Final report submitted to New Leaders for New Schools. Washington, DC: Mathematica Policy Research.

[152] Prawda, J. 1993. *Educational decentralization in latin America: Lessons learned*. International Journal of Educational Development, Vol. 13, No. 3, pp. 253-264.

[153] Psacharopoulos, G. 1994. *Returns to investment in education: A global update*. World Development, Vol. 22, No. 9, pp. 1325–1343.

[154] Psacharopoulos, G. 1985. *Returns to education: A further international update and implications*. Journal of Human Resources, Vol. 20, No. 4, pp. 583–604.

[155] Raudenbush, S. W. and Willms, J. D. 1995. *The estimation of school effects*. Journal of Educational and Behavioral Statistics, Vol. 20, No. 4, pp. 307-335.

[156] Raudenbush, S. W. and Bryk, A. S. 2002. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

[157] Reinikka, R. and Svensson, J. 2004. *Local capture: Evidence from a central government transfer program in Uganda*. The Quarterly Journal of Economics, Vol. 119, No. 2, pp. 679–705.

[158] Rivkin, S. G., Hanushek, E. A. and Kain, J. F. 2005. *Teachers, schools, and academic achievement*. Econometrica, Vol. 73, No. 2, pp. 417-458.

[159] Robins, J. M. and Rotnitzky, A. 1995. *Semiparametric efficiency in multivariate regression models with missing data*. Journal of the American Statistical Association, Vol. 90, No. 429, pp. 122–129.

[160] Robins, J. M., Rotnitzky, A. and Zhao, L. P. 1995. *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data*. Journal of the American Statistical Association, Vol. 90, No. 429, pp. 106-121.

[161] Romano, J. P. and Wolf, M. 2005. *Stepwise multiple testing as formalized data snooping*. Econometrica, Vol. 73, No. 4, pp. 1237-1282.

[162] Rotz, D., Johnson, M. and Gill, B. 2014. *Value-added models for the Pittsburgh public schools, 2012-13 school year*. Retrieved July 2014 from *http://www.mathematica-mpr.com/~/media/publications/pdfs/education/vam_pittsburgh.pdf*.

[163] Rubin, D. B. and Thomas, N. 1996. *Matching using estimated propensity scores: Relating theory to practice*. Biometrics, Vol. 52, No. 1, pp. 249–264.

[164] Rutter, M., Maughan, B., Mortimore, P., Ouston, J. and Smith, A. 1979. *Fifteen thousand hours*. Wells: Open Books.

[165] Schultz, P. 2002. *Why governments should invest more to educate girls*. World Development, Vol. 30, No. 2, pp. 207–225.

[166] Schultz, P. 1997. *The demand for children in low-income countries*. In M. R. Rosenzweig and O. Stark (Eds.), Handbook of Population and Family Economics, pp. 350-418. Amsterdam: North Holland Press.

[167] Schwartz, A. E., Stiefel, L. and Wiswall, M. 2011. *Do small schools improve performance in large, urban districts? Causal evidence from New York city*. Working Paper No. 04-11. Retrieved September 2013 from *http://steinhardt.nyu.edu/scmsAdmin/media/users/qd217/Working_Paper_04-11.pdf*.

[168] Schwille, J. and Eisemont, D. 1991. *Is grade repetition always wasteful? New data and unanswered questions*. Bridge Research Report Series No. 7.

[169] Sen, A. 1999. *Development as freedom*. Oxford: Oxford University Press.

[170] Sen, A. 1992. *Inequality reexamined*. New-York and Cambridge, MA: Russel Sage Foundation, Harvard University Press.

[171] Senegal, Republic of. 1996. *Law No. 96-06 of March 22*. Local Government Code.

[172] Shadish, W. R., Clark, M. H. and Steiner, P. M. 2008. *Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments*. Journal of the American Statistical Association, Vol. 103, No. 484, pp. 1334–1344.

[173] Shin, Y. and Raudenbush, S. W. 2011. *The causal effect of class size on academic achievement: Multivariate instrumental variable estimators with data missing at random*. Journal of Educational and Behavioral Statistics, Vol. 36, No. 2, pp. 154-185.

[174] Shin, Y. and Raudenbush, S. W. 2010. *A latent cluster-mean approach to the contextual effects model with missing data*. Journal of Educational and Behavioral Statistics, Vol. 35, No. 1, pp. 26–53.

[175] Slate, J. R. and Jones, C. 2005. *Effects of school size: A review of the literature with recommendations*. Essays in Education, Vol. 13, pp. 1-22.

[176] Smith, B. C. 1985. *Decentralization: The territorial dimension of the State*. George Allen and Unwin, London.

[177] Smith, D. T. and DeYoung, A. J. 1988. *Big school vs. small school: Conceptual, empirical, and political perspectives on the reemerging debate*. Journal of Rural and Small Schools, Vol. 2, No. 2, pp. 2–11.

[178] Strang, D. 1987. *The administrative transformation of American education: School district consolidation, 1938–1980.* Administrative Science Quarterly, Vol. 32, No. 3, pp. 352–366.

[179] Strauss, J. and Thomas, D. 1995. *Human resources: Empirical modeling of household and family decisions.* In T. N. Srinivasan and J. Behrman (Eds.), Handbook of Development Economics, Vol. 3A, Amsterdam: North Holland Press, Introduction.

[180] Su, Y.-S. and Cortina, J. 2009. *What do we gain? Combining propensity score methods and multilevel modeling.* APSA 2009 Toronto Meeting Paper.

[181] Sun, H., Creemers, B. P. M. and De Jong, R. 2007. *Contextual factors and effective school improvement.* School Effectiveness and School Improvement, Vol. 18, pp. 93-122.

[182] Texas Education Agency. 1999. *School size and class size in Texas public schools.* Report No. 12. Document No. GE9 600 03. Austin, TX: Texas Education Agency Office of Policy Planning and Research.

[183] Tsiatis, A. A. 2006. *Semiparametric theory and missing data.* New York: Springer.

[184] Vandenbussche, J., Aghion, P. and Meghir, C. 2006. *Growth, distance to frontier and composition of human capital.* Journal of Economic Growth, Vol. 11, No. 2, pp. 97-127.

[185] Van Den Noortgate, W., Opdenakker, M. C. and Onghena, P. 2005. *The effects of ignoring a level in multilevel analysis.* School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, Vol. 16, pp. 281–303.

[186] Van Der Laan, M. and Robins, J. 2003. *Unified methods for censored longitudinal data and causality.* Springer, Berlin.

[187] Van Landeghem, G., De Fraine, B. and Van Damme, J. 2005. *The consequence of ignoring a level of nesting in multilevel analysis: A comment.* Multivariate Behavioral Research, Vol. 40, No. 4, pp. 423–434.

[188] Walberg, H. and Walberg, H. 1994. *Losing local control.* Educational Researcher, Vol. 23, No. 5, pp. 19–26.

[189] Willms, J. D. 1992. *Monitoring school performance: A guide for educators*. London: Falmer Press.

[190] Woessmann, L. 2004. *How equal are educational opportunities? Family background and student achievement in Europe and the United States*. Munich: Cesifo Working Paper No. 1162.

[191] Wooldridge, J. M. 2005. *Violating ignorability of treatment by controlling for too many factors*. Econometric Theory, Vol. 21, No. 5, pp. 1026-1028.

[192] World Bank. 2013. *Projet d'amélioration de la qualité et de l'équité de l'éducation de base*. Report No. 75752-SN. Dakar, Sénégal.

[193] Wyse, E., Keesler, V. and Schneider, B. 2008. *Assessing the effects of small school size on mathematics achievement: A propensity score matching approach*. Teachers College Record, Vol. 110, No. 9, pp. 1879-1900.

Appendix A

Supplement to Chapter 1

Figure A.1: *Location of Schools in the Project*

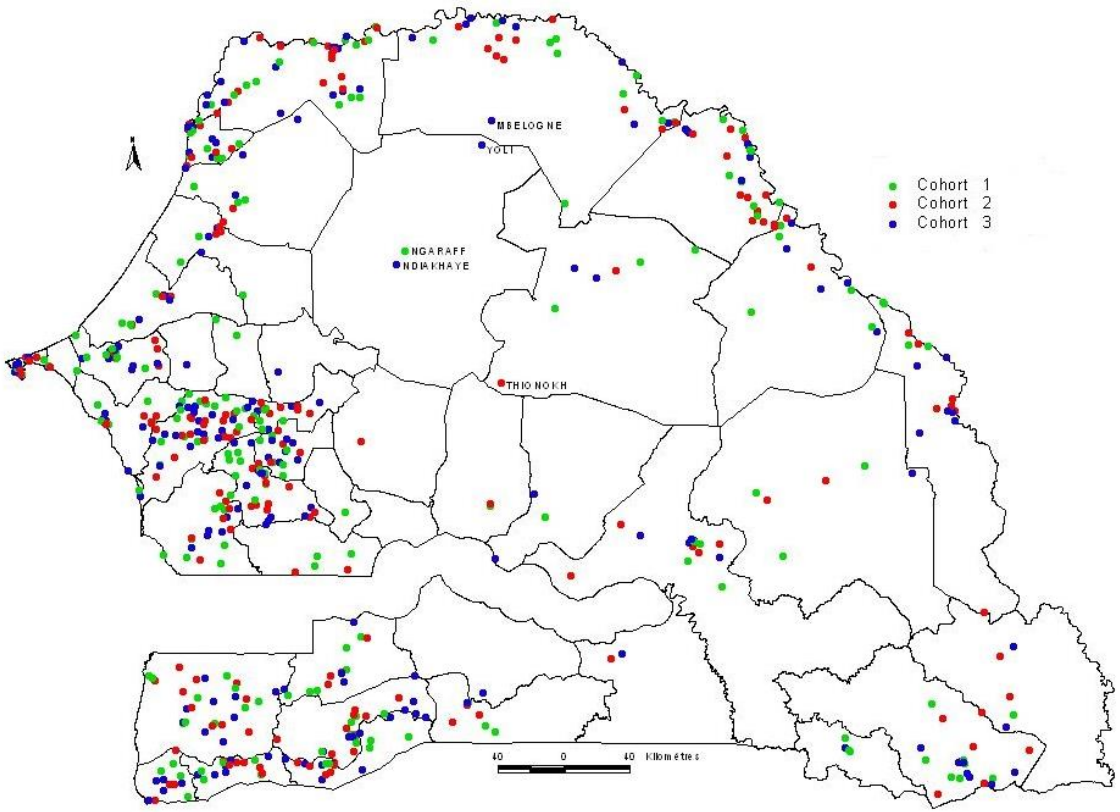


Table A.1: *Summary Statistics (Mean and Standard Deviation) of a Nationally Representative Sample of Second and Fifth Graders*

Panel A: School and Teacher Characteristics

The school locality has a health centre	0.809	(0.393)
The school has electricity	0.359	(0.480)
Number of teachers	9.809	(5.007)
Number of pupils (School size)	500.683	(386.209)
The school has a library	0.217	(0.412)
Percentage of teachers with a Baccalaureate degree	0.474	(0.499)

Panel B: Household Characteristics

The student's father is literate	0.585	(0.493)
The student's mother is literate	0.355	(0.478)
The student's home has electricity	0.595	(0.491)
The student's home has a television	0.598	(0.490)
The student's home has a modern toilet	0.367	(0.482)

Weighted means and standard deviations are shown in parentheses.
Source: Data collected by PASEC in 2007.

Table A.2: *Student Test Scores Sample Sizes and Attrition*

	Grade 3			Grade 5		
	French	Mathematics	Oral	French	Mathematics	Oral
Baseline Sample Size	2,722	2,752	1,388	2,724	2,726	1,362
First Follow-up Sample Size	2,720	2,718	1,385	2,648	2,643	1,347
New Observations	322	299	177	262	261	155
Total Attrition	324	333	180	338	344	170
Percentage of Attrition	0.119	0.121	0.130	0.124	0.126	0.125
Percentage of Attrition, Treated	0.118	0.117	0.113	0.116	0.116	0.114
Percentage of Attrition, Control	0.119	0.123	0.138	0.128	0.131	0.130
Second Follow-up Sample Size	1,732	1,721	853	1,606	1,606	833
Total Attrition*	290	301	208	355	357	186
Percentage of Attrition*	0.157	0.162	0.222	0.197	0.197	0.206
Percentage of Attrition*, Treated	0.160	0.165	0.230	0.204	0.206	0.224
Percentage of Attrition*, Control	0.155	0.159	0.215	0.189	0.189	0.188
Observed in All Waves	1,464	1,461	709	1,396	1,392	696

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

Table A.3: *Baseline Descriptive Statistics and Balance, Grade 4*

Panel A: Test Scores

	Females		Males	
	Control	Differences	Control	Differences
Percentage of correct answers: French	0.387	-0.005	0.396	0.004
	(0.168)	(0.015)	(0.171)	(0.016)
Percentage of correct answers: Mathematics	0.313	-0.001	0.346	0.013
	(0.185)	(0.017)	(0.187)	(0.017)
Percentage of correct answers: Oral	0.537	0.000	0.570	0.016
	(0.241)	(0.024)	(0.244)	(0.022)
Index Score	-0.061	0.029	0.088	0.063
	(0.970)	(0.103)	(0.989)	(0.096)

Panel B: School and Teacher Characteristics

	All Students	
	Control	Differences
Distance of the school locality to the nearest city (in kilometres)	18.033	-0.213
	(24.555)	(2.197)
Locality population (in 100,000s inhabitants)	1.410	-0.031
	(4.428)	(0.453)
The school locality has a health centre	0.714	-0.030
	(0.452)	(0.043)
The school is located in the South of the country	0.187	0.011
	(0.390)	(0.037)
School has electricity	0.566	-0.014
	(0.496)	(0.048)
Number of teachers	9.741	-0.575
	(4.927)	(0.517)
Number of pupils (School size)	343.646	-35.569
	(253.370)	(26.048)
The school has a library	0.210	-0.082*
	(0.408)	(0.043)
Number of computers	1.303	-0.014
	(4.390)	(0.404)
Proportion of female teachers	0.318	-0.010
	(0.234)	(0.023)
Average age of teachers	33.255	0.097
	(4.233)	(0.390)
Percentage of teachers with a Baccalaureate degree	0.413	0.024
	(0.223)	(0.022)
Average experience of teachers	6.607	-0.134
	(3.694)	(0.354)
Teachers had training in the five years preceding the grant	0.468	-0.007
	(0.499)	(0.049)
Number of manuals in classroom	66.431	-5.679
	(51.957)	(5.405)

Table A.3: *Continued*

Panel C: Household Characteristics

	Females		Males	
	Control	Differences	Control	Differences
Absenteeism (in days missed the week preceding the survey)	0.155	0.095*	0.153	0.051
	(0.731)	(0.055)	(0.774)	(0.057)
Proportion of students that participate in housework	0.018	0.010	0.023	0.017
	(0.132)	(0.011)	(0.149)	(0.010)
Household size	9.148	-0.363	9.113	0.107
	(3.951)	(0.466)	(4.172)	(0.402)
Number of children in the household	5.155	-0.522*	5.194	0.081
	(2.671)	(0.300)	(2.883)	(0.284)
Proportion of Heads of household with any education	0.574	0.127**	0.541	0.008
	(0.495)	(0.054)	(0.499)	(0.050)
Proportion of adult females with any education	0.313	0.012	0.322	0.013
	(0.396)	(0.041)	(0.392)	(0.039)
Proportion of literate heads of household	0.590	0.098*	0.584	0.027
	(0.493)	(0.053)	(0.494)	(0.051)
Proportion of literate adult females	0.346	0.037	0.339	-0.007
	(0.397)	(0.041)	(0.391)	(0.039)
Distance of the student's home to school (in kilometres)	0.646	-0.052	0.628	-0.132
	(0.745)	(0.098)	(0.768)	(0.082)
The student's parents are involved in the school activities	0.438	0.092*	0.453	0.051
	(0.497)	(0.052)	(0.499)	(0.049)
Household food expenditures (in 1,000s CFA Francs)	22.679	0.190	21.689	-1.270
	(16.038)	(1.686)	(15.390)	(1.459)
Expenditures on uniform (in 1,000s CFA Francs)	2.360	-0.178	2.259	0.065
	(1.141)	(0.482)	(1.138)	(0.380)
Expenditures on tuition (in 1,000s CFA Francs)	1.060	-0.078	1.034	0.008
	(1.036)	(0.124)	(1.006)	(0.090)
Expenditures on supplies (in 1,000s CFA Francs)	4.328	0.182	4.387	0.573*
	(4.824)	(0.385)	(3.560)	(0.337)
The student has a tutor	0.148	0.066**	0.139	-0.044
	(0.356)	(0.033)	(0.347)	(0.039)
The student's home has electricity	0.451	-0.015	0.455	-0.019
	(0.498)	(0.054)	(0.499)	(0.053)
The student's home has a television	0.482	0.039	0.490	0.047
	(0.501)	(0.054)	(0.501)	(0.051)
The student's home has a modern toilet	0.493	-0.011	0.529	0.002
	(0.501)	(0.056)	(0.500)	(0.051)
Land owned by the household (in hectares)	3.200	0.585	2.605	0.434
	(12.592)	(0.815)	(4.270)	(0.383)
The interview is conducted in French language	0.088	0.013	0.129	0.005
	(0.284)	(0.028)	(0.336)	(0.033)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4.

Table A.4: *First Follow-up Descriptive Statistics and Balance, Grade 3*

Panel A: School Characteristics

	All Students	
	Control	Differences
Distance of the school locality to the nearest city (in kilometres)	18.347	-0.097
	(24.781)	(2.104)
Locality population (in 100,000s inhabitants)	0.917	-0.269
	(2.744)	(0.313)
The school locality has a health centre	0.709	-0.026
	(0.454)	(0.043)
The school is located in the South of the country	0.195	-0.005
	(0.396)	(0.037)

Panel B: Household Characteristics

	Females		Males	
	Control	Differences	Control	Differences
Household size	10.026	-0.127	9.891	-0.135
	(4.586)	(0.466)	(4.264)	(0.440)
Number of children in the household	5.481	-0.354	5.459	-0.112
	(2.852)	(0.299)	(2.773)	(0.273)
Heads of household with any education	0.398	-0.025	0.445	0.010
	(0.490)	(0.048)	(0.498)	(0.050)
Adult females with any education	0.232	0.020	0.216	-0.037
	(0.347)	(0.034)	(0.329)	(0.037)
Distance of home to school (in kilometres)	0.640	0.120**	0.593	0.064
	(0.698)	(0.059)	(0.564)	(0.055)
The student's home has a television	0.461	-0.023	0.447	-0.053
	(0.499)	(0.052)	(0.498)	(0.051)
The student's home has a modern toilet	0.355	-0.027	0.372	0.060
	(0.479)	(0.050)	(0.484)	(0.047)
Land owned by the household (in hectares)	2.403	-0.872	2.795	-0.190
	(4.043)	(0.541)	(5.213)	(0.587)
The interview is conducted in French language	0.090	-0.018	0.116	-0.001
	(0.287)	(0.031)	(0.320)	(0.031)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard deviations in parentheses in columns 1 and 3. Clustered standard errors in parentheses in columns 2 and 4.

Table A.5: *Differences in Baseline Characteristics among First Follow-up Non-attriters*

Panel A: Test Scores

	Grade 2		Grade 4	
	Females	Males	Females	Males
Percentage of correct answers: French	-0.003	0.013	-0.010	0.009
	(0.019)	(0.021)	(0.016)	(0.015)
Percentage of correct answers: Mathematics	0.001	0.009	-0.012	0.016
	(0.019)	(0.020)	(0.018)	(0.017)
Percentage of correct answers: Oral	-0.019	-0.007	-0.001	0.012
	(0.017)	(0.019)	(0.025)	(0.022)
Index Score	-0.028	0.056	0.014	0.068
	(0.105)	(0.113)	(0.100)	(0.094)

Panel B: School and Teacher Characteristics

	Grade 2	Grade 4
Distance of the school to the nearest city (in kilometres)	-0.347	-0.035
	(2.203)	(2.341)
Locality population (in 100,000s inhabitants)	0.063	-0.142
	(0.407)	(0.473)
The school locality has a health centre	-0.032	-0.032
	(0.044)	(0.043)
The school is located in the South of the country	0.016	0.021
	(0.038)	(0.037)
The school has electricity	-0.021	-0.037
	(0.049)	(0.049)
Number of teachers	-0.464	-0.534
	(0.515)	(0.515)
Number of pupils (School size)	-33.657	-31.181
	(25.444)	(26.044)
The school has a library	-0.085**	-0.094**
	(0.043)	(0.044)
Number of computers in the school	0.000	-0.052
	(0.397)	(0.409)
Percentage of female teachers in the school	-0.007	-0.012
	(0.023)	(0.023)
Average age of teachers	0.126	-0.051
	(0.394)	(0.392)
Percentage of teachers with a Baccalaureate degree	0.012	0.017
	(0.023)	(0.023)
Average experience of teachers	-0.085	-0.220
	(0.345)	(0.355)
Teachers had training in the five years preceding the grant	-0.092*	-0.012
	(0.050)	(0.050)
Number of manuals in the classroom	-3.810	-6.530
	(4.564)	(5.498)

Table A.5: *Continued*

Panel C: Household characteristics

	Grade 2		Grade 4	
	Females	Males	Females	Males
Absenteeism (in days missed the week preceding the survey)	-0.074	-0.078	0.100*	0.053
	(0.114)	(0.092)	(0.058)	(0.063)
Proportion of students that participate in housework	-0.017	0.000	0.012	0.015
	(0.016)	(0.009)	(0.012)	(0.011)
Household size	-0.069	0.358	-0.439	0.045
	(0.432)	(0.474)	(0.512)	(0.419)
Number of children in the household	-0.045	0.276	-0.579*	0.074
	(0.281)	(0.315)	(0.329)	(0.304)
Proportion of heads of household with any education	0.095*	-0.046	0.132**	0.035
	(0.053)	(0.054)	(0.058)	(0.053)
Proportion of adult females with any education	0.020	0.004	-0.006	0.026
	(0.044)	(0.046)	(0.043)	(0.042)
Proportion of literate heads of household	0.094*	-0.015	0.104*	0.023
	(0.053)	(0.054)	(0.057)	(0.053)
Proportion of literate adult females	0.004	-0.009	0.014	-0.003
	(0.043)	(0.045)	(0.045)	(0.041)
Distance of the student's home to school (in kilometres)	0.051	0.131	-0.062	-0.165*
	(0.084)	(0.094)	(0.108)	(0.092)
The student's parents are involved in the school activities	-0.114**	0.009	0.070	0.070
	(0.053)	(0.053)	(0.055)	(0.052)
Household food expenditures (in 1,000s CFA Francs)	-2.122	0.121	-0.324	-2.192
	(1.670)	(1.572)	(1.850)	(1.568)
Expenditures on uniform (in 1,000s CFA Francs)	-0.543	0.271	-0.064	0.313
	(0.528)	(0.537)	(0.507)	(0.368)
Expenditures on tuition (in 1,000s CFA Francs)	-0.066	0.011	-0.016	-0.001
	(0.111)	(0.128)	(0.123)	(0.097)
Expenditures on supplies (in 1,000s CFA Francs)	0.047	0.986*	0.011	0.693*
	(0.254)	(0.565)	(0.416)	(0.371)
The student has a tutor	-0.008	0.049	0.055	-0.060
	(0.038)	(0.037)	(0.035)	(0.042)
The student's home has electricity	-0.067	-0.016	-0.081	-0.024
	(0.055)	(0.055)	(0.057)	(0.055)
The student's home has a television	-0.056	-0.053	-0.020	0.046
	(0.054)	(0.056)	(0.057)	(0.054)
The student's home has a modern toilet	-0.020	0.047	-0.045	0.017
	(0.055)	(0.054)	(0.059)	(0.054)
Land owned by the household (in hectares)	0.045	-0.661	0.120	0.310
	(0.469)	(0.915)	(0.466)	(0.421)
The interview is conducted in French language	0.026	0.048	0.002	0.025
	(0.030)	(0.033)	(0.031)	(0.034)

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

Table A.6: *Differences in Baseline Characteristics among Second Follow-up Non-attriters*

Panel A: Test Scores

	Grade 2		Grade 4	
	Females	Males	Females	Males
Percentage of correct answers: French	-0.021	0.009	-0.004	0.001
	(0.023)	(0.025)	(0.018)	(0.018)
Percentage of correct answers: Mathematics	-0.011	0.011	-0.004	0.011
	(0.022)	(0.024)	(0.020)	(0.020)
Percentage of correct answers: Oral	-0.020	-0.001	0.006	0.004
	(0.019)	(0.023)	(0.029)	(0.027)
Index Score	-0.104	-0.008	0.049	0.005
	(0.128)	(0.133)	(0.116)	(0.110)

Panel B: School and Teacher Characteristics

	Grade 2	Grade 4
Distance of the school locality to the nearest city (in kilometres)	0.513	0.985
	(2.436)	(2.579)
Locality population (in 100,000s inhabitants)	-0.160	-0.295
	(0.442)	(0.500)
The school locality has a health centre	-0.076	-0.069
	(0.052)	(0.052)
The school is located in the South of the country	0.024	0.040
	(0.046)	(0.044)
The school has electricity	0.019	-0.003
	(0.056)	(0.057)
Number of teachers	-0.636	-0.708
	(0.587)	(0.609)
Number of pupils (School size)	-43.355	-47.491
	(28.603)	(30.330)
The school has a library	-0.087*	-0.079
	(0.047)	(0.049)
Number of computers	0.144	0.170
	(0.485)	(0.515)
Proportion of female teachers	-0.028	-0.032
	(0.026)	(0.026)
Average age of teachers	0.056	-0.020
	(0.459)	(0.454)
Percentage of teachers with a Baccalaureate degree	0.016	0.017
	(0.027)	(0.027)
Average experience of teachers	-0.138	-0.168
	(0.394)	(0.404)
Teachers had training in the five years preceding the grant	-0.071	-0.024
	(0.058)	(0.058)
Number of manuals in classroom	-3.572	-10.830*
	(5.758)	(6.374)

Table A.6: *Continued*

Panel C: Household Characteristics

	Grade 2		Grade 4	
	Females	Males	Females	Males
Absenteeism (in days missed the week preceding the survey)	-0.043	-0.064	0.126*	0.152
	(0.169)	(0.111)	(0.076)	(0.097)
Proportion of students that participate in housework	-0.026	0.007	0.015	0.031
	(0.018)	(0.013)	(0.017)	(0.019)
Household size	-0.551	-0.339	-0.426	0.399
	(0.513)	(0.538)	(0.596)	(0.529)
Number of children in the household	-0.207	0.027	-0.466	0.249
	(0.341)	(0.366)	(0.404)	(0.375)
Proportion of heads of household with any education	0.074	-0.041	0.105	0.036
	(0.064)	(0.063)	(0.069)	(0.064)
Proportion of adult females with any education	0.025	-0.017	-0.038	0.035
	(0.055)	(0.055)	(0.053)	(0.049)
Proportion of literate heads of household	0.040	0.009	0.077	0.004
	(0.066)	(0.063)	(0.067)	(0.064)
Proportion of literate adult females	0.004	-0.050	-0.008	0.018
	(0.054)	(0.053)	(0.055)	(0.050)
Distance of the student's home to school (in kilometres)	0.097	0.163	-0.045	-0.206**
	(0.116)	(0.103)	(0.117)	(0.102)
The student's parents are involved in the school activities	-0.097	-0.021	0.041	0.087
	(0.065)	(0.063)	(0.066)	(0.063)
Household food expenditures (in 1,000s CFA Francs)	-3.075	-1.924	0.047	-0.751
	(2.020)	(1.853)	(2.281)	(2.110)
Expenditures on uniform (in 1,000s CFA Francs)	-0.297	0.438	0.122	0.500
	(0.736)	(0.644)	(0.693)	(0.442)
Expenditures on tuition (in 1,000s CFA Francs)	-0.053	0.129	0.092	0.126
	(0.135)	(0.177)	(0.170)	(0.146)
Expenditures on supplies (in 1,000s CFA Francs)	-0.129	1.238	0.650	0.425
	(0.355)	(0.982)	(0.670)	(0.445)
The student has a tutor	-0.020	0.028	0.064	-0.049
	(0.041)	(0.043)	(0.043)	(0.046)
The student's home has electricity	-0.070	0.004	-0.067	-0.019
	(0.067)	(0.066)	(0.070)	(0.068)
The student's home has a television	-0.085	-0.084	0.011	0.074
	(0.066)	(0.066)	(0.069)	(0.066)
The student's home has a modern toilet	0.052	0.006	-0.054	0.065
	(0.066)	(0.064)	(0.069)	(0.066)
Land owned by the household (in hectares)	-0.257	-1.246	-0.206	0.481
	(0.550)	(1.009)	(0.462)	(0.575)
The interview is conducted in French language	0.020	0.038	0.006	-0.010
	(0.039)	(0.041)	(0.036)	(0.041)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Point estimates of the difference in characteristic between non-treated students and treated students, among students who did not attrit between baseline and second follow-up. Student in the second follow-up have attrited if they have a baseline test score but not a second follow-up test score (regardless of their status in the first follow-up).
Clustered standard errors in parentheses.

Table A.7: *Characteristics of Schools by Baseline Missing Status, First Follow-up*

Panel A: Treatment Status

	Not Missing	Missing	Differences
Treated	0.337	0.312	0.025
	(0.022)	(0.086)	(0.089)

Panel B: Control School Characteristics

	Not Missing	Missing	Differences
The school is located in the South of the country	0.186	0.439	-0.253**
	(0.022)	(0.114)	(0.116)
The school is located in a rural area	0.736	1.000	-0.264***
	(0.025)	(0.000)	(0.025)
Locality population (in 100,000s inhabitants)	0.981	0.024	0.957***
	(0.173)	(0.006)	(0.173)
Number of teachers	9.946	7.090	2.856***
	(0.278)	(0.951)	(0.991)
Number of pupils (School size)	341.617	238.326	103.291***
	(14.283)	(35.026)	(37.826)
Percentage of correct answers: French, Grade 3	0.540	0.418	0.123***
	(0.010)	(0.044)	(0.045)
Percentage of correct answers: Mathematics, Grade 3	0.550	0.408	0.142***
	(0.009)	(0.041)	(0.042)
Percentage of correct answers: Oral, Grade 3	0.361	0.212	0.149***
	(0.010)	(0.036)	(0.037)
Percentage of correct answers: Index, Grade 3	-0.012	-0.666	0.653***
	(0.047)	(0.197)	(0.203)
Percentage of correct answers: French, Grade 5	0.483	0.379	0.104***
	(0.008)	(0.030)	(0.032)
Percentage of correct answers: Mathematics, Grade 5	0.441	0.342	0.099***
	(0.009)	(0.036)	(0.037)
Percentage of correct answers: Oral, Grade 5	0.648	0.572	0.076
	(0.011)	(0.049)	(0.050)
Percentage of correct answers: Index, Grade 5	0.026	-0.483	0.509***
	(0.045)	(0.173)	(0.179)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Clustered standard errors in parentheses.

Table A.8: *Programme Impacts on Grade 3 Test Scores, Standardized Coefficients*

Panel A: Beginning of Third Grade, First Follow-up

	French	Mathematics	Oral
Females	0.149	0.128	0.210
	[0.022, 0.272]	[0.013, 0.247]	[0.061, 0.356]
Males	0.088	0.098	0.050
	[-0.053, 0.209]	[-0.029, 0.218]	[-0.102, 0.194]

Panel B: End of Third Grade, Second Follow-up

	French	Mathematics	Oral
Females	0.182	0.083	0.228
	[0.030, 0.344]	[-0.076, 0.236]	[0.041, 0.411]
Males	0.111	0.068	0.102
	[-0.032, 0.264]	[-0.086, 0.205]	[-0.081, 0.277]

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

Table A.9: *Programme Impacts on Grade 3 Test Scores, No Conditioning Variables*

Panel A: Beginning of Third Grade, First Follow-up

	French	Mathematics	Oral	Index[a]
Females	0.041	0.031	0.055	0.240
	[0.004, 0.081]	[-0.009, 0.066]	[0.014, 0.097]	[0.054, 0.415]
Males	0.018	0.020	0.012	0.042
	[-0.019, 0.056]	[-0.017, 0.057]	[-0.029, 0.053]	[-0.137, 0.231]

Panel B: End of Third Grade, Second Follow-up

	French	Mathematics	Oral	Index[a]	PPVT[a]
Females	0.046	0.021	0.065	0.290	0.134
	[-0.000, 0.091]	[-0.019, 0.062]	[0.013, 0.116]	[0.061, 0.529]	[-0.137, 0.386]
Males	0.020	0.008	0.025	0.068	0.198
	[-0.024, 0.061]	[-0.031, 0.047]	[-0.024, 0.076]	[-0.134, 0.280]	[-0.032, 0.436]

[a] The index and PPVT columns are in standardized units.
Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering.

Table A.10: *Programme Impacts on Grade 5 Test Scores*

Panel A: Beginning of Fifth Grade, First Follow-up

	French	Mathematics	Oral	Index[a]
Females	0.009	0.013	-0.005	0.009
	[-0.014, 0.033]	[-0.017, 0.041]	[-0.034, 0.025]	[-0.104, 0.131]
Males	0.014	0.008	0.020	0.043
	[-0.009, 0.035]	[-0.021, 0.034]	[-0.011, 0.051]	[-0.084, 0.171]

Panel B: End of Fifth Grade, Second Follow-up

	French	Mathematics	Oral	Index[a]	PPVT[a]
Females	0.008	-0.003	-0.006	-0.022	-0.096
	[-0.019, 0.033]	[-0.029, 0.023]	[-0.046, 0.030]	[-0.163, 0.108]	[-0.339, 0.155]
Males	-0.001	-0.013	0.019	0.046	-0.029
	[-0.029, 0.027]	[-0.044, 0.013]	[-0.019, 0.055]	[-0.113, 0.212]	[-0.280, 0.218]

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

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

Table A.11: *Programme Impacts on Test Scores at Different Grades*

Panel A: Beginning of Grade, First Follow-up

	French		Mathematics		Oral		Index[a]	
Grade	Females	Males	Females	Males	Females	Males	Females	Males
3	0.041**	0.018	0.031*	0.020	0.055***	0.012	0.240**	0.042
	(0.020)	(0.019)	(0.018)	(0.018)	(0.021)	(0.021)	(0.096)	(0.095)
	[0.244]	[0.249]	[0.237]	[0.238]	[0.221]	[0.220]	[0.992]	[0.968]
5	0.019	0.004	0.023	-0.003	-0.001	0.010	0.030	0.008
	(0.017)	(0.016)	(0.017)	(0.017)	(0.022)	(0.022)	(0.097)	(0.097)
	[0.201]	[0.203]	[0.198]	[0.207]	[0.235]	[0.240]	[0.956]	[1.018]
6	0.016	0.016	0.024	0.014				
	(0.016)	(0.015)	(0.016)	(0.016)				
	[0.195]	[0.189]	[0.201]	[0.196]				

Panel B: End of Grade, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
Grade	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
2	-0.010	0.013	-0.005	-0.000						
	(0.023)	(0.023)	(0.022)	(0.021)						
	[0.254]	[0.260]	[0.246]	[0.241]						
3	0.046**	0.020	0.021	0.008	0.065**	0.025	0.290	0.068	0.134	0.198*
	(0.022)	(0.021)	(0.020)	(0.021)	(0.027)	(0.026)	(0.118)	(0.109)	(0.131)	(0.117)
	[0.237]	[0.231]	[0.233]	[0.221]	[0.230]	[0.223]	[1.026]	[0.917]	[1.006]	[0.949]
4	-0.009	0.001	-0.021	0.003						
	(0.019)	(0.017)	(0.019)	(0.019)						
	[0.201]	[0.195]	[0.201]	[0.208]						
5	0.009	-0.005	-0.002	-0.020	-0.005	0.020	-0.046	0.051	-0.106	-0.046
	(0.020)	(0.018)	(0.020)	(0.019)	(0.024)	(0.021)	(0.115)	(0.109)	(0.128)	(0.125)
	[0.195]	[0.198]	[0.206]	[0.202]	[0.212]	[0.199]	[0.953]	[0.988]	[1.006]	[1.022]
6	0.004	0.007	0.000	-0.002						
	(0.017)	(0.016)	(0.018)	(0.016)						
	[0.183]	[0.174]	[0.193]	[0.169]						

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[a] The index and PPVT columns are in standardized units.
Clustered standard errors in parentheses and standard deviation of control sample in brackets.

Figure A.2: *Distribution of Second-Third Grades French Scores by Gender*

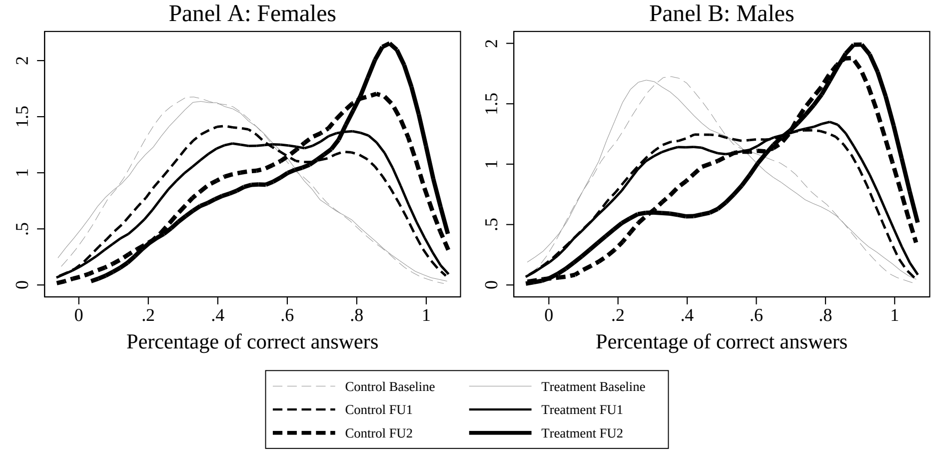


Figure A.3: *Distribution of Second-Third Grades Mathematics Scores by Gender*

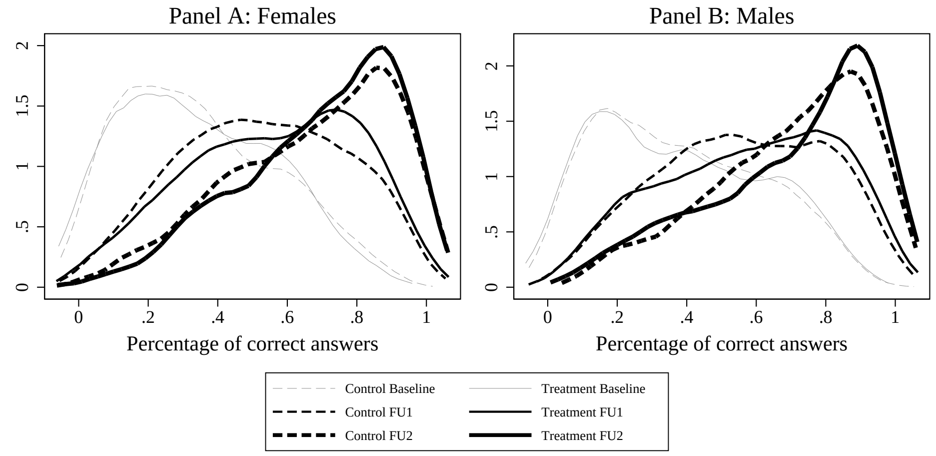


Figure A.4: *Distribution of Second-Third Grades Oral Scores by Gender*

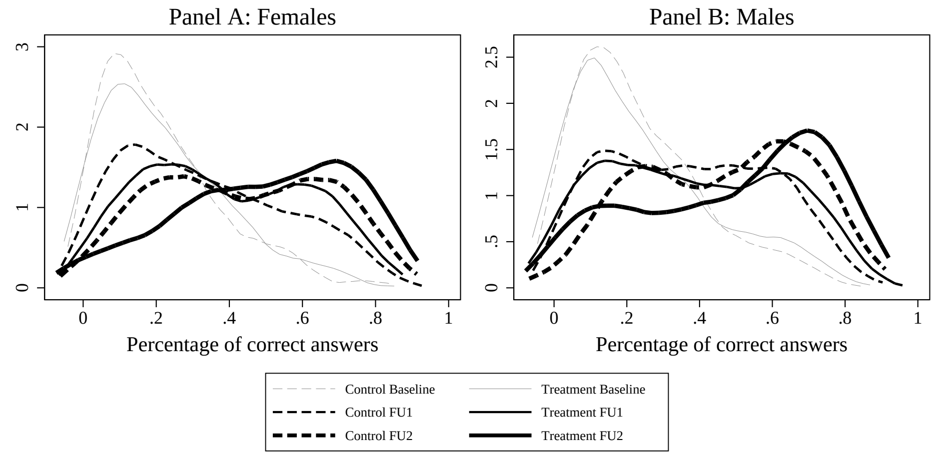


Figure A.5: *Distribution of Fourth-Fifth Grades French Scores by Gender*

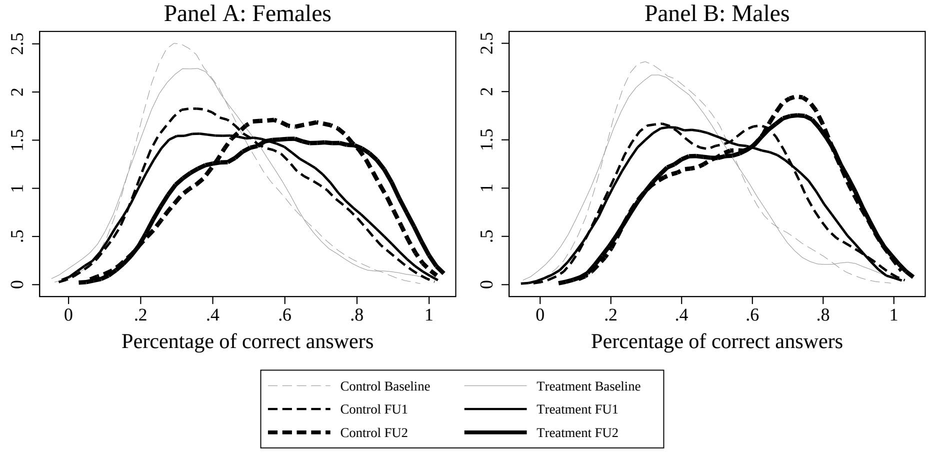


Figure A.6: *Distribution of Fourth-Fifth Grades Mathematics Scores by Gender*

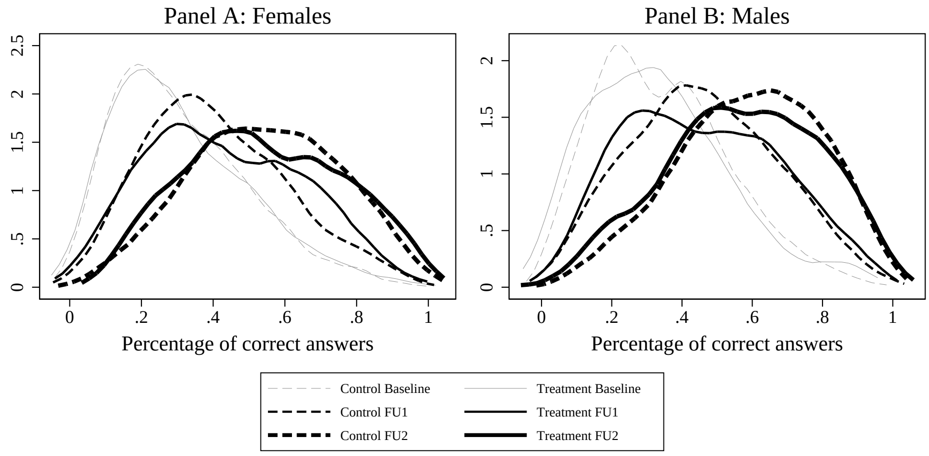


Figure A.7: *Distribution of Fourth-Fifth Grades Oral Scores by Gender*

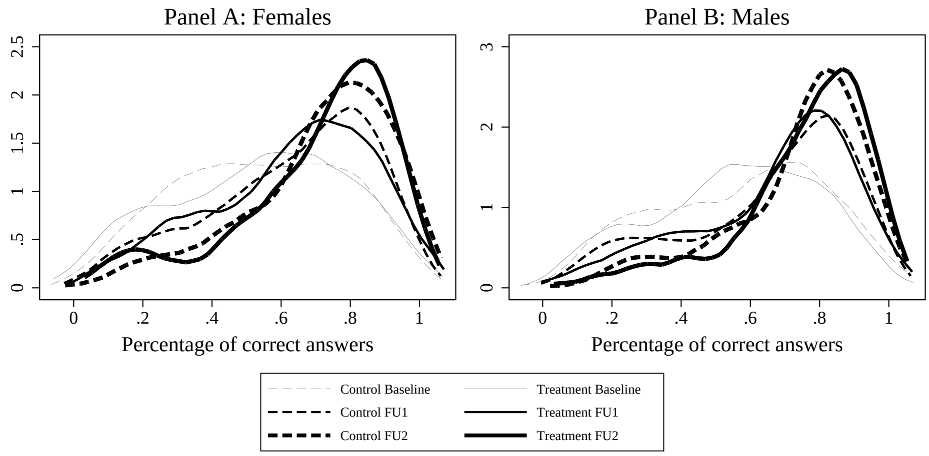
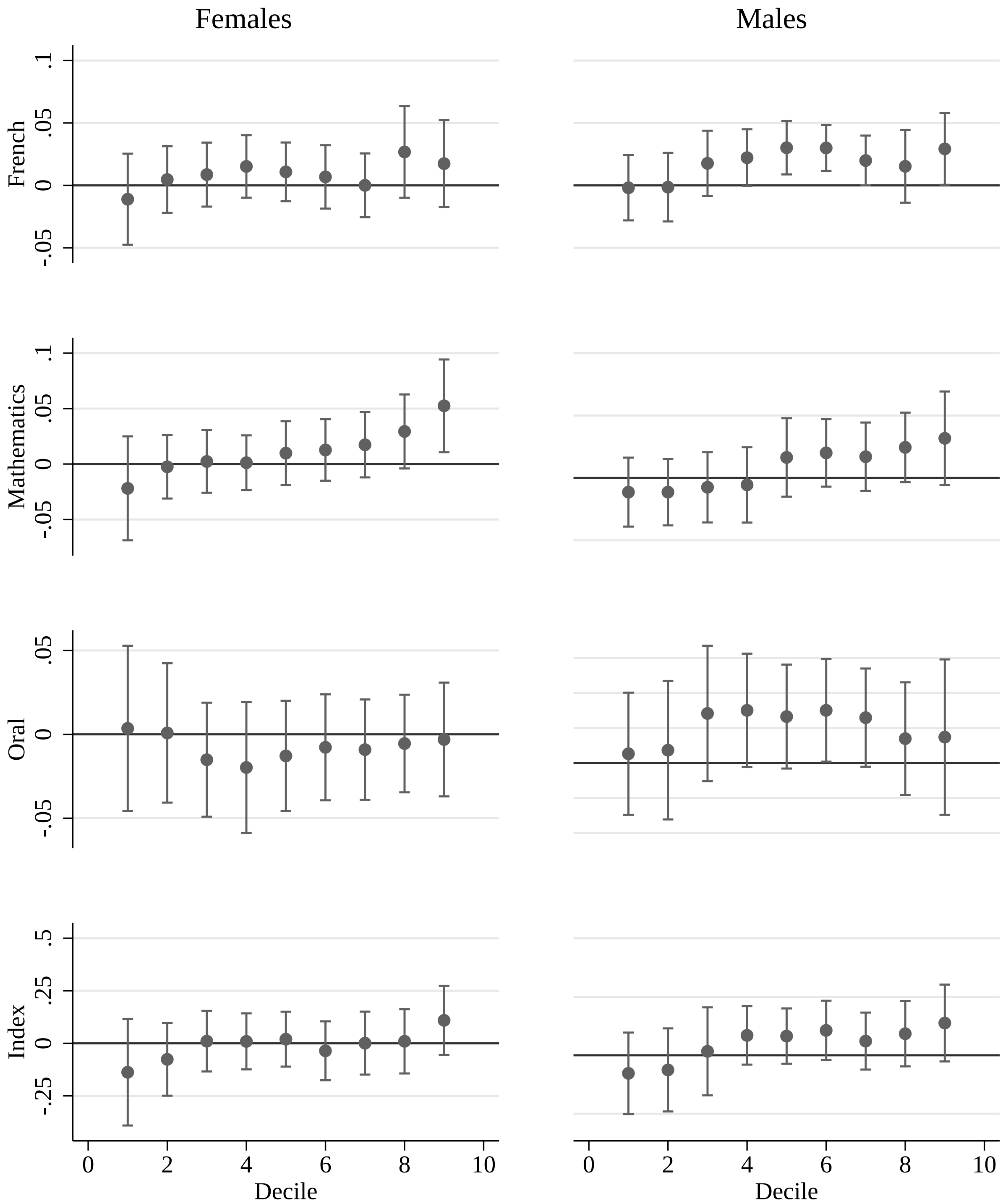
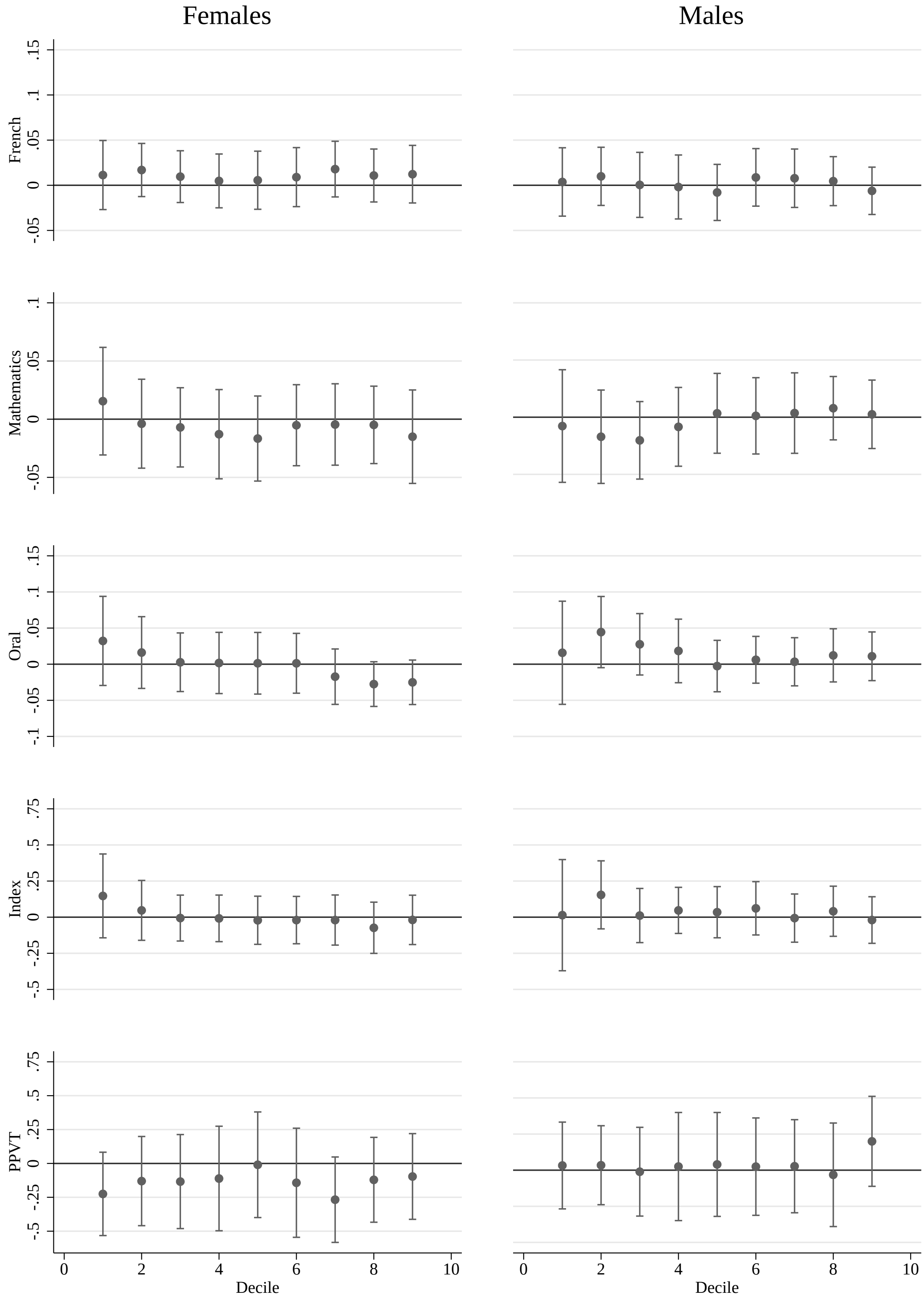


Figure A.8: *Distributional Impacts on Test Scores at the Beginning of Fifth Grade, First Follow-up*



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

Figure A.9: *Distributional Impacts on Test Scores at the End of Fifth Grade, Second Follow-up*



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

Table A.12: *Programme Impacts on Grade 3 Test Scores by Region, Standardized Coefficients*

Panel A: Beginning of Grade 3, First Follow-up

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

Panel B: End of Grade 3, Second Follow-up

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for control variables.

Table A.13: *Programme Impacts on Grade 5 Test Scores by Baseline Ability*

Panel A: Beginning of Grade 5, First Follow-up

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

Panel B: End of Grade 5, Second Follow-up

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

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.14: *Programme Impacts on Grade 5 Test Scores by Region*

Panel A: Beginning of Grade 5, First Follow-up

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

Panel B: End of Grade 5, Second Follow-up

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

[a] The index and PPVT columns are in standardized units.
Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.15: *Programme Impacts on Grade 3 Test Scores by School Size*

Panel A: Beginning of Grade 3, First Follow-up

	French		Mathematics		Oral		Index[a]	
	Females	Males	Females	Males	Females	Males	Females	Males
Low	0.023	0.019	0.013	0.003	0.030	0.004	0.143	0.028
	[-0.024, 0.068]	[-0.023, 0.061]	[-0.027, 0.052]	[-0.036, 0.042]	[-0.019, 0.075]	[-0.039, 0.049]	[-0.076, 0.314]	[-0.174, 0.212]
High	0.032	0.007	0.034	0.022	0.048	-0.004	0.192	-0.050
	[-0.013, 0.077]	[-0.045, 0.057]	[-0.006, 0.075]	[-0.017, 0.061]	[0.002, 0.095]	[-0.050, 0.036]	[-0.001, 0.392]	[-0.271, 0.138]
p-value	0.781	0.715	0.447	0.507	0.582	0.806	0.735	0.577

Panel B: End of Grade 3, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Low	0.020	0.010	-0.003	0.001	0.037	-0.014	0.120	0.032	0.171	0.230
	[-0.031, 0.069]	[-0.039, 0.061]	[-0.046, 0.042]	[-0.049, 0.045]	[-0.026, 0.105]	[-0.068, 0.040]	[-0.132, 0.371]	[-0.184, 0.268]	[-0.140, 0.463]	[-0.088, 0.545]
High	0.057	0.024	0.035	0.016	0.041	0.049	0.262	0.080	0.107	0.098
	[-0.007, 0.118]	[-0.028, 0.082]	[-0.013, 0.087]	[-0.034, 0.065]	[-0.034, 0.100]	[-0.013, 0.116]	[-0.013, 0.542]	[-0.157, 0.350]	[-0.278, 0.484]	[-0.240, 0.419]
p-value	0.361	0.701	0.250	0.665	0.937	0.151	0.484	0.775	0.802	0.582

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.16: *Programme Impacts on Grade 5 Test Scores by School Size*

Panel A: Beginning of Grade 5, First Follow-up

	French		Mathematics		Oral		Index[a]	
	Females	Males	Females	Males	Females	Males	Females	Males
Low	0.022	0.031	0.030	0.029	0.018	0.036	0.091	0.122
	[-0.012, 0.054]	[0.002, 0.059]	[-0.002, 0.063]	[-0.001, 0.062]	[-0.019, 0.056]	[-0.003, 0.076]	[-0.065, 0.239]	[-0.033, 0.274]
High	-0.013	-0.009	-0.013	-0.025	-0.031	0.013	-0.093	-0.006
	[-0.051, 0.024]	[-0.045, 0.025]	[-0.052, 0.023]	[-0.066, 0.012]	[-0.074, 0.016]	[-0.032, 0.058]	[-0.269, 0.090]	[-0.201, 0.188]
p-value	0.166	0.089	0.093	0.037	0.093	0.453	0.126	0.331

Panel B: End of Grade 5, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Low	0.016	-0.003	0.022	-0.002	0.026	-0.001	0.131	-0.033	-0.069	-0.162
	[-0.018, 0.050]	[-0.037, 0.029]	[-0.017, 0.056]	[-0.037, 0.036]	[-0.022, 0.070]	[-0.047, 0.043]	[-0.066, 0.324]	[-0.243, 0.166]	[-0.441, 0.284]	[-0.489, 0.161]
High	-0.002	-0.006	-0.029	-0.035	-0.041	0.031	-0.167	0.081	-0.257	0.062
	[-0.040, 0.037]	[-0.052, 0.041]	[-0.067, 0.010]	[-0.086, 0.012]	[-0.097, 0.011]	[-0.025, 0.083]	[-0.384, 0.051]	[-0.184, 0.342]	[-0.606, 0.107]	[-0.341, 0.424]
p-value	0.508	0.906	0.065	0.288	0.069	0.362	0.038	0.498	0.463	0.372

[a] The index and PPVT columns are in standardized units.
Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.17: *Programme Impacts on Grade 3 Test Scores by Wealth Status*

Panel A: Beginning of Grade 3, First Follow-up

	French		Mathematics		Oral		Index[a]	
	Females	Males	Females	Males	Females	Males	Females	Males
Low	0.015	0.047	0.030	0.031	0.031	0.023	0.115	0.170
	[-0.048, 0.076]	[-0.014, 0.111]	[-0.025, 0.085]	[-0.023, 0.086]	[-0.026, 0.088]	[-0.040, 0.080]	[-0.135, 0.357]	[-0.104, 0.419]
High	0.053	-0.010	0.058	0.001	0.032	0.027	0.164	-0.008
	[0.006, 0.104]	[-0.070, 0.057]	[0.013, 0.104]	[-0.050, 0.051]	[-0.018, 0.085]	[-0.027, 0.078]	[-0.032, 0.376]	[-0.267, 0.240]
p-value	0.355	0.198	0.456	0.430	0.983	0.918	0.760	0.319

Panel B: End of Grade 3, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Low	0.034	0.023	0.048	0.002	0.058	0.029	0.306	0.124	0.191	0.065
	[-0.039, 0.114]	[-0.053, 0.096]	[-0.026, 0.119]	[-0.066, 0.061]	[-0.027, 0.141]	[-0.061, 0.117]	[-0.022, 0.658]	[-0.216, 0.498]	[-0.188, 0.602]	[-0.307, 0.391]
High	0.061	0.085	0.021	0.056	0.049	0.077	0.120	0.356	0.142	0.180
	[-0.005, 0.135]	[0.023, 0.152]	[-0.041, 0.082]	[0.002, 0.111]	[-0.017, 0.118]	[0.005, 0.151]	[-0.203, 0.452]	[0.036, 0.647]	[-0.228, 0.483]	[-0.107, 0.488]
p-value	0.605	0.239	0.555	0.225	0.878	0.434	0.432	0.308	0.853	0.643

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.18: *Programme Impacts on Grade 5 Test Scores by Wealth Status*

Panel A: Beginning of Grade 5, First Follow-up

	French		Mathematics		Oral		Index[a]	
	Females	Males	Females	Males	Females	Males	Females	Males
Low	-0.010	0.014	0.013	0.018	-0.037	0.011	-0.095	0.072
	[-0.053, 0.034]	[-0.024, 0.050]	[-0.038, 0.064]	[-0.023, 0.061]	[-0.091, 0.012]	[-0.039, 0.056]	[-0.330, 0.126]	[-0.145, 0.267]
High	-0.007	0.026	-0.016	0.016	-0.005	0.026	-0.046	0.080
	[-0.044, 0.029]	[-0.013, 0.068]	[-0.058, 0.026]	[-0.024, 0.055]	[-0.049, 0.043]	[-0.017, 0.067]	[-0.206, 0.120]	[-0.111, 0.263]
p-value	0.901	0.671	0.387	0.931	0.330	0.648	0.728	0.955

Panel B: End of Grade 5, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Low	-0.017	-0.021	-0.010	-0.010	-0.036	0.051	-0.027	0.126	-0.128	-0.127
	[-0.068, 0.035]	[-0.078, 0.029]	[-0.063, 0.051]	[-0.071, 0.046]	[-0.108, 0.027]	[-0.018, 0.126]	[-0.316, 0.265]	[-0.172, 0.413]	[-0.494, 0.314]	[-0.522, 0.275]
High	0.022	0.005	-0.004	-0.016	0.020	-0.021	0.018	-0.118	-0.206	0.084
	[-0.023, 0.069]	[-0.043, 0.051]	[-0.058, 0.047]	[-0.067, 0.036]	[-0.044, 0.084]	[-0.072, 0.022]	[-0.175, 0.272]	[-0.343, 0.115]	[-0.520, 0.131]	[-0.226, 0.395]
p-value	0.292	0.460	0.873	0.884	0.197	0.100	0.819	0.219	0.770	0.402

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.19: *Programme Impacts on Grade 3 Test Scores by Urban and Rural Location of Schools*

Panel A: Beginning of Grade 3, First Follow-up

	French		Mathematics		Oral		Index[a]	
	Females	Males	Females	Males	Females	Males	Females	Males
Rural	0.021	0.023	0.025	0.032	0.046	0.021	0.161	0.078
	[-0.013, 0.061]	[-0.021, 0.060]	[-0.010, 0.059]	[-0.005, 0.067]	[0.009, 0.086]	[-0.022, 0.061]	[-0.026, 0.304]	[-0.115, 0.252]
Urban	0.043	0.003	0.020	-0.014	0.021	-0.021	0.199	-0.100
	[-0.007, 0.096]	[-0.045, 0.056]	[-0.024, 0.064]	[-0.054, 0.028]	[-0.031, 0.077]	[-0.074, 0.029]	[-0.014, 0.424]	[-0.308, 0.097]
p-value	0.494	0.572	0.858	0.111	0.479	0.214	0.794	0.197

Panel B: End of Grade 3, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Rural	0.019	0.018	0.009	0.004	0.041	0.016	0.212	0.023	0.018	0.243
	[-0.034, 0.062]	[-0.029, 0.070]	[-0.035, 0.054]	[-0.039, 0.044]	[-0.012, 0.092]	[-0.040, 0.069]	[-0.028, 0.429]	[-0.202, 0.232]	[-0.316, 0.314]	[-0.048, 0.490]
Urban	0.087	0.024	0.027	0.025	0.021	0.023	0.140	0.162	0.294	0.173
	[0.035, 0.149]	[-0.036, 0.087]	[-0.033, 0.081]	[-0.028, 0.078]	[-0.051, 0.102]	[-0.037, 0.088]	[-0.192, 0.456]	[-0.087, 0.397]	[-0.046, 0.706]	[-0.155, 0.497]
p-value	0.072	0.895	0.641	0.538	0.678	0.875	0.736	0.418	0.269	0.750

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.20: *Programme Impacts on Grade 5 Test Scores by Urban and Rural Location of Schools*

Panel A: Beginning of Grade 5, First Follow-up

	French		Mathematics		Oral		Index[a]	
	Females	Males	Females	Males	Females	Males	Females	Males
Rural	0.014	0.015	0.021	0.006	0.005	0.020	0.055	0.068
	[-0.019, 0.045]	[-0.017, 0.043]	[-0.012, 0.055]	[-0.028, 0.037]	[-0.029, 0.039]	[-0.019, 0.059]	[-0.097, 0.203]	[-0.094, 0.227]
Urban	-0.017	0.000	-0.022	-0.002	-0.038	0.025	-0.137	0.004
	[-0.056, 0.021]	[-0.033, 0.032]	[-0.064, 0.020]	[-0.038, 0.035]	[-0.083, 0.012]	[-0.019, 0.071]	[-0.335, 0.052]	[-0.178, 0.192]
p-value	0.227	0.500	0.116	0.766	0.153	0.848	0.124	0.598

Panel B: End of Grade 5, Second Follow-up

	French		Mathematics		Oral		Index[a]		PPVT[a]	
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Rural	0.006	-0.010	-0.008	-0.023	0.005	0.025	0.015	0.072	-0.067	0.116
	[-0.027, 0.040]	[-0.045, 0.027]	[-0.048, 0.028]	[-0.064, 0.012]	[-0.040, 0.050]	[-0.013, 0.068]	[-0.154, 0.185]	[-0.127, 0.273]	[-0.360, 0.239]	[-0.182, 0.421]
Urban	-0.000	0.006	0.002	-0.000	-0.038	0.015	-0.108	-0.001	-0.332	-0.420
	[-0.039, 0.036]	[-0.041, 0.052]	[-0.043, 0.046]	[-0.043, 0.044]	[-0.105, 0.026]	[-0.050, 0.077]	[-0.353, 0.143]	[-0.288, 0.262]	[-0.749, 0.083]	[-0.846, 0.069]
p-value	0.804	0.586	0.728	0.451	0.300	0.786	0.403	0.670	0.324	0.064

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

Bootstrapped confidence intervals in square brackets are calculated using the 2.5 and 97.5 percentiles of the bootstrap distribution and are adjusted for clustering. See text or table 1.2 for conditioning variables. The p-value row shows the p-value of the test of equality of the two treatment effects of the column.

Table A.21: *Programme Impacts on Teacher and Household Outcomes, First Follow-up*

Panel A: Teacher Outcomes

	Grade 3	Grade 5	Differences
The teacher has a Baccalaureate degree	-0.052	0.018	-0.070
	(0.047)	(0.046)	(0.064)
Teachers had training in the five years preceding the grant	0.083**	0.101**	-0.018
	(0.039)	(0.047)	(0.056)
Number of minutes spent preparing lessons	2.115	0.061	2.054
	(1.726)	(1.614)	(2.023)
Number of manuals	6.209	6.493	-0.284
	(4.231)	(4.964)	(5.296)
Number of measuring instruments	0.138	0.228	-0.090
	(0.179)	(0.193)	(0.210)
Number of chairs	0.018	0.071**	-0.053
	(0.038)	(0.035)	(0.041)
The teacher use books while teaching	0.004	0.012	-0.008
	(0.013)	(0.012)	(0.015)
The teacher use computers while teaching	-0.005	0.034	-0.038
	(0.019)	(0.024)	(0.029)
Times per day the teachers have to ask for silence	-1.638**	1.109	-2.747***
	(0.686)	(0.830)	(0.931)
Times per day the teachers have to punish a Student	-0.126	0.741**	-0.868**
	(0.277)	(0.370)	(0.407)
Students leaving the school in the year preceding the grant	0.303	1.063	-0.760
	(0.730)	(0.800)	(1.095)
Students joining the school in the year preceding the grant	0.262	-0.030	0.292
	(0.212)	(0.185)	(0.194)

Panel B: Household Outcomes

	Grade 3	Grade 5	Differences
Absenteeism (in days missed the week preceding the survey)	0.106	0.010	0.096
	(0.075)	(0.055)	(0.086)
Proportion of students that participate in housework	-0.008	-0.016	0.008
	(0.012)	(0.014)	(0.014)
Student's parents are involved in school activities	0.037	0.026	0.011
	(0.038)	(0.038)	(0.039)
Expenditures on uniform (in 1,000s CFA Francs)	-0.003	-0.070	0.067
	(0.070)	(0.056)	(0.055)
Expenditures on tuition (in 1,000s CFA Francs)	0.303	0.035	0.269
	(0.303)	(0.164)	(0.313)
Expenditures on supplies (in 1,000s CFA Francs)	-0.162	-0.111	-0.050
	(0.232)	(0.320)	(0.302)
The student has a tutor	-0.027	-0.009	-0.017
	(0.026)	(0.029)	(0.031)
Expenditures on children (1,000s CFA Francs)	0.303	-0.175	0.477
	(0.611)	(0.662)	(0.683)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Clustered standard errors in parentheses.

Table A.22: *Programme Impacts on Grade 3 Household Characteristics by Region, First Follow-up*

	Females			Males		
	South	North	Differences	South	North	Differences
Proportion of students that participate in housework	-0.016	-0.033*	0.017	0.047	0.006	0.041
	(0.016)	(0.019)	(0.025)	(0.050)	(0.019)	(0.053)
Absenteeism (in days missed the week preceding the survey)	0.498*	0.107	0.391	0.195	0.009	0.185
	(0.277)	(0.087)	(0.291)	(0.167)	(0.133)	(0.214)
The student's parents are involved in the school activities	0.026	0.043	-0.017	-0.094	0.061	-0.155
	(0.102)	(0.058)	(0.117)	(0.101)	(0.055)	(0.115)
Expenditures on uniform (in 1,000s CFA Francs)	0.011	0.026	-0.015	-0.034	-0.039	0.005
	(0.240)	(0.083)	(0.254)	(0.335)	(0.054)	(0.339)
Expenditures on tuition (in 1,000s CFA Francs)	1.294	0.064	1.230	1.404	0.059	1.345
	(1.418)	(0.158)	(1.427)	(1.509)	(0.132)	(1.515)
Expenditures on supplies (in 1,000s CFA Francs)	-0.042	-0.026	-0.015	-0.040	-0.221	0.181
	(0.532)	(0.356)	(0.640)	(0.561)	(0.342)	(0.658)
The student has a tutor	-0.032	-0.046	0.014	-0.044	-0.013	-0.031
	(0.022)	(0.042)	(0.048)	(0.049)	(0.042)	(0.064)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Clustered standard errors in parentheses.

Appendix B

Supplement to Chapter 2

Table B.1: *Mean and Standard Deviation of Locality- and School-level Covariates by Urban and Rural Location of Schools*

	Rural	Urban	Total
Distance from the school to the country's capital (Dakar) (in kilometres)	355.725	234.465	323.339
	(208.230)	(201.372)	(213.239)
Distance of the school to the nearest urban centre (in kilometres)	21.316	10.287	18.370
	(21.111)	(28.325)	(23.758)
Number of inhabitants in the city/village of the school	11,351.381	464,231.213	132,305.624
	(48,091.351)	(769,966.970)	(447,219.975)
There is a health centre in the locality of the school	0.642	0.924	0.717
	(0.479)	(0.265)	(0.450)
Age of the school's principal (in years)	41.974	51.413	44.495
	(7.825)	(6.933)	(8.668)
The school's principal is a female	0.010	0.051	0.021
	(0.098)	(0.220)	(0.142)
The school's principal has a BFEM or less	0.331	0.234	0.305
	(0.471)	(0.424)	(0.461)
Experience of the school's principal (in years)	17.340	28.244	20.253
	(8.022)	(7.964)	(9.347)
The school was selected in 2009 for a school grant project	0.310	0.421	0.340
	(0.463)	(0.494)	(0.474)
The school is entirely or partially electrified	0.455	0.837	0.557
	(0.498)	(0.370)	(0.497)
The school has piped water	0.499	0.820	0.585
	(0.500)	(0.384)	(0.493)
The school has a library	0.167	0.399	0.229
	(0.373)	(0.490)	(0.420)

Table B.2: *Mean and Standard Deviation of Other Covariates, Grade 2*

	Rural	Urban	Total
Age of the teacher (in years)	31.385 (6.417)	35.592 (8.247)	32.518 (7.202)
The teacher is a female	0.435 (0.496)	0.685 (0.465)	0.502 (0.500)
The teacher has a Baccalaureate/BSEN degree	0.354 (0.478)	0.311 (0.463)	0.342 (0.475)
The teaching language is French	0.839 (0.367)	0.912 (0.283)	0.859 (0.348)
Classroom resources: number of French manuals	28.192 (18.934)	35.964 (22.521)	30.285 (20.254)
Classroom resources: number of mathematics manuals	22.908 (20.237)	33.285 (24.830)	25.702 (22.051)
The class (grade) size is greater than 40	0.357 (0.479)	0.641 (0.480)	0.433 (0.496)
The student is a female	0.503 (0.500)	0.493 (0.500)	0.500 (0.500)
Age of the student (in years)	8.834 (1.045)	8.789 (1.042)	8.822 (1.044)
Time the student spends daily on housework (in hours)	1.337 (1.732)	0.676 (1.145)	1.153 (1.617)
Student's household size	9.680 (4.379)	8.498 (3.486)	9.350 (4.181)
The student goes to school regularly	0.905 (0.294)	0.836 (0.371)	0.885 (0.319)
The student has a chronic disability	0.065 (0.247)	0.078 (0.268)	0.069 (0.253)
The student works after school to earn money	0.014 (0.118)	0.009 (0.095)	0.013 (0.112)
The student has help with school work at home	0.081 (0.273)	0.320 (0.467)	0.148 (0.355)
Distance of the student's home to his/her school (in kilometres)	0.839 (1.192)	0.750 (0.897)	0.814 (1.118)
The student attends the closest school to his/her home	0.931 (0.254)	0.740 (0.440)	0.878 (0.328)
The student walks to school	0.952 (0.213)	0.950 (0.219)	0.952 (0.215)
Travel time from student's home to school (in minutes)	11.502 (11.837)	12.098 (14.313)	11.668 (12.570)
The student lives with his/her biological parents	0.735 (0.442)	0.717 (0.452)	0.730 (0.444)
The head of household is a female	0.065 (0.247)	0.142 (0.349)	0.087 (0.281)
Age of the head of household (in years)	48.939 (12.111)	49.242 (11.964)	49.024 (12.064)
The head of household has no formal education	0.440 (0.497)	0.384 (0.487)	0.424 (0.495)
The head of household attended koranic school or has completed primary school	0.438 (0.497)	0.384 (0.487)	0.423 (0.494)
The head of household has at least a primary education level	0.122 (0.327)	0.233 (0.424)	0.153 (0.360)
The head of household is literate	0.578 (0.494)	0.685 (0.466)	0.608 (0.489)
Socioeconomic status of the student's household	2.631 (1.830)	4.653 (2.066)	3.195 (2.103)
The head of household is jobless	0.087 (0.281)	0.169 (0.376)	0.110 (0.313)
The head of household works in the public sector	0.074 (0.262)	0.187 (0.391)	0.106 (0.308)
The head of household works in the formal sector or the informal private sector	0.800 (0.400)	0.589 (0.493)	0.741 (0.438)
The head of household is retired	0.039 (0.193)	0.055 (0.228)	0.043 (0.204)
The head of household is monogamous	0.530 (0.500)	0.594 (0.492)	0.548 (0.498)
Proportion of correct answers on the French test	0.325 (0.211)	0.412 (0.226)	0.348 (0.218)
Proportion of correct answers on the mathematics test	0.273 (0.208)	0.374 (0.226)	0.300 (0.217)

Table B.3: *Mean and Standard Deviation of Other Covariates, Grade 4*

	Rural	Urban	Total
Age of the teacher (in years)	32.011 (6.073)	36.087 (8.219)	33.100 (6.949)
The teacher is a female	0.298 (0.457)	0.479 (0.500)	0.346 (0.476)
The teacher has a Baccalaureate/BSEN degree	0.346 (0.476)	0.480 (0.500)	0.382 (0.486)
The teaching language is French	0.906 (0.292)	0.926 (0.262)	0.911 (0.285)
Classroom resources: number of French manuals	24.896 (16.823)	35.665 (25.152)	27.772 (19.971)
Classroom resources: number of mathematics manuals	23.402 (18.325)	29.715 (20.785)	25.088 (19.213)
The class (grade) size is greater than 40	0.280 (0.449)	0.543 (0.499)	0.350 (0.477)
The student is a female	0.491 (0.500)	0.479 (0.500)	0.487 (0.500)
Age of the student (in years)	10.951 (1.205)	11.035 (1.319)	10.973 (1.236)
Time the student spends daily on housework (in hours)	1.630 (1.802)	1.005 (1.240)	1.461 (1.691)
Student's household size	9.751 (4.225)	8.274 (3.467)	9.351 (4.085)
The student goes to school regularly	0.908 (0.289)	0.856 (0.352)	0.894 (0.308)
The student has a chronic disability	0.057 (0.232)	0.047 (0.211)	0.054 (0.226)
The student works after school to earn money	0.017 (0.130)	0.014 (0.118)	0.016 (0.127)
The student has help with school work at home	0.069 (0.254)	0.307 (0.462)	0.134 (0.340)
Distance of the student's home to his/her school (in kilometres)	0.758 (0.973)	0.927 (1.323)	0.804 (1.081)
The student attends the closest school to his/her home	0.931 (0.254)	0.763 (0.426)	0.885 (0.319)
The student walks to school	0.965 (0.183)	0.935 (0.247)	0.957 (0.203)
Travel time from student's home to school (in minutes)	11.768 (12.007)	11.530 (10.791)	11.703 (11.684)
The student lives with his/her biological parents	0.741 (0.439)	0.744 (0.437)	0.742 (0.438)
The head of household is a female	0.076 (0.265)	0.163 (0.370)	0.099 (0.300)
Age of the head of household (in years)	50.649 (12.447)	49.435 (10.952)	50.320 (12.066)
The head of household has no formal education	0.504 (0.500)	0.395 (0.490)	0.475 (0.500)
The head of household attended koranic school or has completed primary school	0.389 (0.488)	0.316 (0.466)	0.369 (0.483)
The head of household has at least a primary education level	0.107 (0.309)	0.288 (0.454)	0.156 (0.363)
The head of household is literate	0.499 (0.500)	0.693 (0.462)	0.552 (0.498)
Socioeconomic status of the student's household	2.732 (2.086)	5.443 (2.308)	3.466 (2.462)
The head of household is jobless	0.090 (0.286)	0.191 (0.394)	0.117 (0.322)
The head of household works in the public sector	0.057 (0.232)	0.163 (0.370)	0.086 (0.280)
The head of household works in the formal sector or the informal private sector	0.807 (0.395)	0.581 (0.494)	0.746 (0.436)
The head of household is retired	0.047 (0.211)	0.065 (0.247)	0.052 (0.221)
The head of household is monogamous	0.509 (0.500)	0.581 (0.494)	0.529 (0.499)
Proportion of correct answers on the French test	0.351 (0.171)	0.418 (0.187)	0.369 (0.178)
Proportion of correct answers on the mathematics test	0.294 (0.182)	0.347 (0.191)	0.308 (0.186)

Table B.4: *Mean and Standard Deviation of Characteristics of Small and Large Schools*

	Small Schools	Large Schools	Total	T-statistics of the difference
The school in located in a rural region	0.859 (0.348)	0.421 (0.496)	0.746 (0.436)	-8.99
Distance to Dakar (in kilometres)	351.193 (212.346)	292.851 (221.244)	336.141 (215.956)	-2.53
Distance to the nearest urban centre (in kilometres)	22.371 (25.311)	10.628 (16.920)	19.341 (23.977)	-5.73
Locality population (in 1,000s inhabitants)	48.213 (255.183)	323.262 (678.205)	119.939 (426.530)	4.36
There is a health centre in the locality of the school	0.615 (0.487)	0.942 (0.234)	0.699 (0.459)	9.71
The school is entirely or partially electrified	0.471 (0.500)	0.744 (0.438)	0.542 (0.499)	5.68
The school has piped water	0.474 (0.500)	0.785 (0.412)	0.554 (0.498)	6.75
Age of the school's oldest buildings (in years)	21.161 (19.423)	36.860 (21.567)	25.211 (21.125)	7.08
Age of the school's newest buildings (in years)	5.848 (11.269)	7.843 (12.675)	6.362 (11.667)	1.54
Number of teachers in the school	7.204 (3.264)	14.926 (3.362)	9.196 (4.715)	21.96
Total number of students in the school (School size)	207.534 (99.009)	585.537 (114.673)	305.058 (195.069)	32.36
Number of classrooms in the school	6.152 (2.358)	11.653 (2.040)	7.571 (3.316)	24.54
The school got a grant within the past five years	0.414 (0.493)	0.636 (0.483)	0.471 (0.500)	4.35
The school has at least one multi-grade class	0.241 (0.429)	0.017 (0.128)	0.183 (0.387)	-8.73
The school has a library	0.178 (0.383)	0.331 (0.472)	0.217 (0.413)	3.21
If library, number of French manuals for ten students	8.299 (4.904)	7.832 (10.098)	8.121 (7.279)	-0.27
If library, number of mathematics manuals for ten students	4.071 (3.635)	3.567 (4.802)	3.883 (4.090)	-0.54
Number of tables in the school	99.261 (63.987)	219.339 (73.410)	130.241 (84.757)	16.03
Age of the tables (in years)	10.118 (7.246)	12.425 (7.130)	10.711 (7.279)	3.05
Number of computers in the school	0.552 (2.834)	2.620 (5.662)	1.085 (3.872)	3.86
Number of toilets for teachers	0.782 (0.986)	1.223 (1.242)	0.896 (1.074)	3.55
Number of toilets for students	3.032 (3.238)	5.785 (4.099)	3.742 (3.679)	6.71
The head teacher participates in the decision-making process	0.997 (0.054)	1.000 (0.000)	0.998 (0.046)	1.00
The teachers participate in the decision-making process	1.000 (0.000)	0.983 (0.128)	0.996 (0.065)	-1.42
The school management committee participates in the decision-making process	0.925 (0.263)	0.950 (0.218)	0.932 (0.252)	1.03
The parents participate in the decision-making process	0.948 (0.222)	0.917 (0.276)	0.940 (0.237)	-1.11
The school keeps attendance record	0.905 (0.293)	0.901 (0.300)	0.904 (0.295)	-0.14
Age of the head teacher (in years)	41.410 (8.330)	51.117 (5.924)	43.910 (8.861)	13.84
The head teacher is a female	0.014 (0.119)	0.041 (0.200)	0.021 (0.145)	1.40
The head teacher has at least a Baccalaureate/BSEN degree	0.689 (0.463)	0.760 (0.429)	0.707 (0.456)	1.55
The head teacher has a BFEM or less	0.310 (0.463)	0.240 (0.429)	0.292 (0.455)	-1.53
Experience of the head teacher (in years)	16.725 (8.810)	27.562 (6.593)	19.533 (9.552)	14.20
Class Size - Grade 2	34.360 (14.273)	56.317 (21.046)	40.331 (19.071)	10.58
Class Size - Grade 4	31.581 (14.050)	48.864 (15.402)	36.164 (16.303)	10.70

Table B.5: *Attrition Models for Midline and Endline Surveys by Grade*

	Grade 2		Grade 4	
	Model 1	Model 2	Model 1	Model 2
	Student in Round 1 but not in Round 2	Student in Round 1 but not in Round 3	Student in Round 1 but not in Round 2	Student in Round 1 but not in Round 3
Treatment (School with a large size)	-0.043	-0.001	0.013	0.045
	(0.032)	(0.045)	(0.029)	(0.050)
Distance from the school to the country's capital (Dakar) (in kilometres)	0.000*	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
The school is located in an urban area	0.032	-0.003	0.056*	-0.040
	(0.031)	(0.046)	(0.033)	(0.055)
Distance of the school to the nearest urban centre (in kilometres)	0.000	0.001	-0.001**	-0.001*
	(0.001)	(0.001)	(0.000)	(0.001)
Number of inhabitants in the city/village of the school	0.000	0.000	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
There is a health centre in the locality of the school	0.020	0.040	-0.038	0.022
	(0.024)	(0.039)	(0.027)	(0.039)
Age of the head teacher (in years)	-0.002	-0.009*	0.000	-0.004
	(0.004)	(0.005)	(0.004)	(0.005)
The head teacher is a female	0.080	0.148	-0.043	-0.130***
	(0.104)	(0.136)	(0.069)	(0.045)
The head teacher has a BFEM or less	0.007	0.010	0.042*	-0.042
	(0.027)	(0.037)	(0.026)	(0.037)
Experience of the head teacher (in years)	0.005	0.007	0.001	0.006
	(0.004)	(0.005)	(0.004)	(0.005)
The school was selected in 2009 for a school grant project	0.000	0.059*	0.017	0.016
	(0.024)	(0.031)	(0.023)	(0.034)
Age of the teacher (in years)	-0.001	0.003	0.001	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)
The teacher is a female	0.007	-0.033	-0.008	0.025
	(0.023)	(0.032)	(0.024)	(0.041)
The teacher has a Baccalaureate/BSEN degree	0.010	0.000	-0.002	-0.012
	(0.023)	(0.034)	(0.023)	(0.037)
The teaching language is French	0.002	0.015	-0.003	-0.002
	(0.027)	(0.039)	(0.039)	(0.066)

Table B.5: *Continued*

	Grade 2		Grade 4	
	Model 1	Model 2	Model 1	Model 2
	Student in Round 1 but not in Round 2	Student in Round 1 but not in Round 3	Student in Round 1 but not in Round 2	Student in Round 1 but not in Round 3
The student is a female	-0.012	-0.010	0.036*	-0.018
	(0.019)	(0.027)	(0.021)	(0.027)
Age of the student (in years)	-0.056***	-0.041***	-0.059***	-0.051***
	(0.015)	(0.015)	(0.011)	(0.016)
Time the student spends daily on housework (in hours)	-0.002	0.005	-0.011*	-0.010
	(0.007)	(0.009)	(0.006)	(0.009)
Student's household size	-0.003	-0.003	-0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.004)
The student goes to school regularly	0.005	-0.043	-0.004	-0.015
	(0.031)	(0.050)	(0.030)	(0.049)
The student has a chronic disability	-0.054	0.020	0.079	0.147*
	(0.035)	(0.063)	(0.053)	(0.081)
The student works after school to earn money	-0.013	-0.053	-0.049	-0.159***
	(0.033)	(0.050)	(0.081)	(0.058)
The student has help with school work at home	0.044	0.152***	-0.019	0.071
	(0.034)	(0.055)	(0.035)	(0.054)
Distance from the student's home to school (in kilometres)	0.024**	0.029	-0.019*	-0.015
	(0.012)	(0.018)	(0.011)	(0.019)
The student attends the closest school to his/her home	-0.053	-0.092	-0.016	-0.050
	(0.045)	(0.067)	(0.039)	(0.064)
The student walks to school	0.119***	0.038	-0.091	-0.061
	(0.043)	(0.087)	(0.066)	(0.091)
Travel time from student's home to school (in minutes)	0.000	0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
The student lives with his/her biological parents	-0.026	-0.007	-0.070***	-0.069*
	(0.026)	(0.033)	(0.026)	(0.040)
The head of household is a female	-0.009	0.008	-0.017	-0.080
	(0.045)	(0.061)	(0.036)	(0.062)
Age of the head of household (in years)	-0.000	-0.002	-0.001	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
The head of household has no formal education	0.062	0.039	-0.018	0.013
	(0.041)	(0.053)	(0.044)	(0.065)

Table B.5: *Continued*

	Grade 2		Grade 4	
	Model 1	Model 2	Model 1	Model 2
	Student in Round 1 but not in Round 2	Student in Round 1 but not in Round 3	Student in Round 1 but not in Round 2	Student in Round 1 but not in Round 3
The head of household attended koranic school or has completed primary school	0.046	-0.004	-0.037	0.006
	(0.032)	(0.043)	(0.037)	(0.045)
The head of household is literate	-0.021	-0.008	-0.002	-0.045
	(0.032)	(0.045)	(0.028)	(0.049)
Socioeconomic status of the student's household	-0.004	0.001	-0.004	-0.020**
	(0.007)	(0.008)	(0.006)	(0.008)
The head of household is jobless	0.067	0.036	0.044	-0.040
	(0.055)	(0.079)	(0.047)	(0.077)
The head of household works in the public sector	0.016	-0.065	0.043	-0.008
	(0.052)	(0.070)	(0.057)	(0.089)
The head of household works in the formal sector or the informal private sector	0.048	-0.024	0.078*	-0.054
	(0.048)	(0.065)	(0.043)	(0.069)
The head of household is monogamous	-0.034	0.023	0.009	-0.063**
	(0.022)	(0.029)	(0.023)	(0.031)
Standardized value of the prior French test	-0.010	-0.034*	0.000	-0.024
	(0.013)	(0.018)	(0.018)	(0.027)
Standardized value of the prior mathematics test	-0.001	0.004	-0.006	0.001
	(0.014)	(0.019)	(0.018)	(0.027)
Intercept	0.495**	0.758***	0.897***	1.422***
	(0.202)	(0.257)	(0.223)	(0.298)
Observations	834	557	842	561
Adjusted R^2	0.047	0.071	0.068	0.039

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

Robust standard errors accounting for clustering in parentheses.

Figure B.1: *Boxplots for the school size variable*

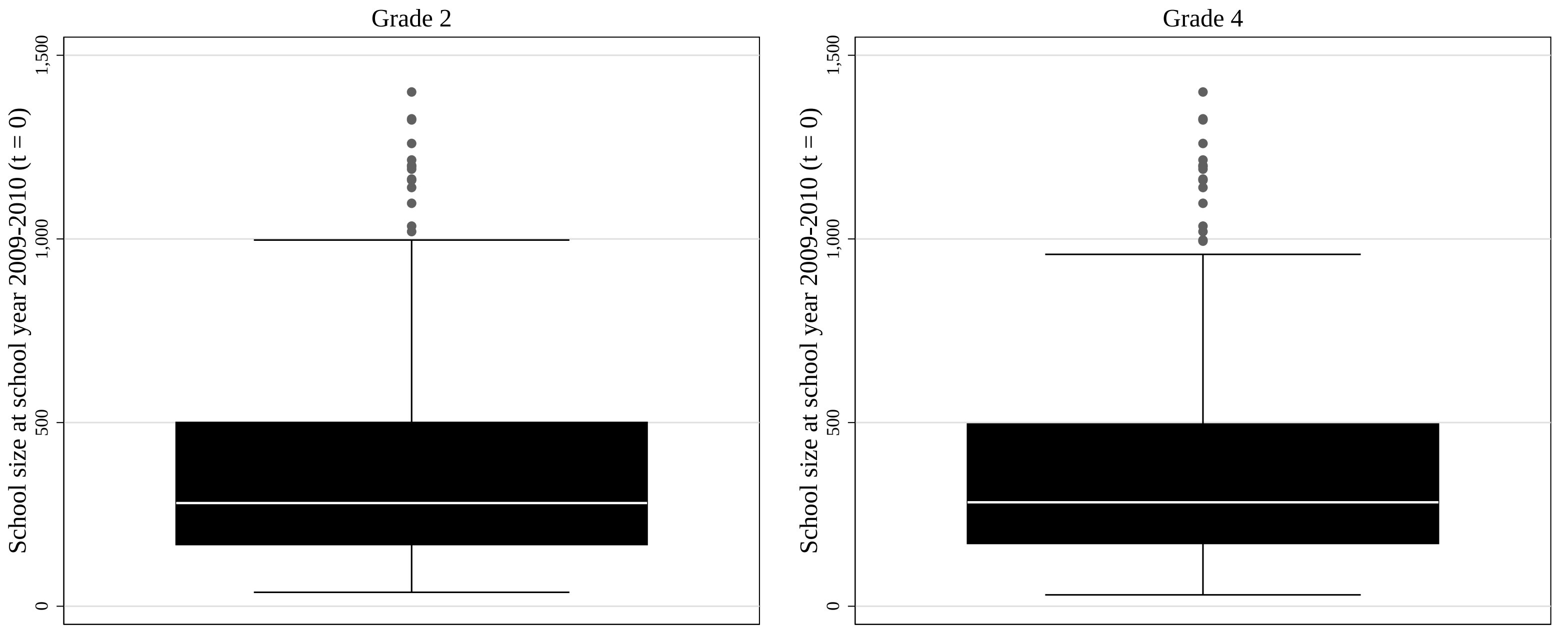


Table B.6: *Estimated Impacts of Large School Size under Different Computation Strategies, Grade 2*

		Set 1 of Covariates			Set 2 of Covariates		
		Naïve difference	Ordinary least squares	Ordinary least squares on the common support	Naïve difference	Ordinary least squares	Ordinary least squares on the common support
French 2010	Estimated Effect	0.221***	-0.022	-0.020	0.221***	-0.006	-0.012
		(0.078)	(0.074)	(0.075)	(0.078)	(0.091)	(0.091)
	Observations	2,209	2,209	2,029	2,209	710	573
Mathematics 2010	Estimated Effect	0.190***	-0.058	-0.053	0.190***	-0.048	-0.026
		(0.073)	(0.057)	(0.057)	(0.073)	(0.077)	(0.078)
	Observations	2,209	2,209	2,029	2,209	710	573
Oral 2010	Estimated Effect	0.240***	-0.017	-0.000	0.240***	0.048	0.058
		(0.090)	(0.076)	(0.076)	(0.090)	(0.094)	(0.096)
	Observations	1,116	1,116	1,022	1,116	615	492
Index 2010	Estimated Effect	0.231**	-0.053	-0.041	0.231**	-0.008	0.006
		(0.092)	(0.072)	(0.072)	(0.092)	(0.087)	(0.087)
	Observations	1,104	1,104	1,012	1,104	603	482
French 2011	Estimated Effect	0.062	-0.117	-0.109	0.062	-0.103	-0.085
		(0.094)	(0.089)	(0.089)	(0.094)	(0.118)	(0.118)
	Observations	1,347	1,347	1,233	1,347	461	373
Mathematics 2011	Estimated Effect	0.111	-0.058	-0.056	0.111	-0.099	-0.080
		(0.092)	(0.082)	(0.081)	(0.092)	(0.113)	(0.112)
	Observations	1,347	1,347	1,233	1,347	461	373
Oral 2011	Estimated Effect	0.227**	-0.026	-0.012	0.227**	0.055	0.073
		(0.112)	(0.099)	(0.099)	(0.112)	(0.123)	(0.118)
	Observations	681	681	619	681	376	301
Index 2011	Estimated Effect	0.140	-0.078	-0.062	0.140	-0.073	-0.049
		(0.120)	(0.102)	(0.101)	(0.120)	(0.129)	(0.123)
	Observations	661	661	600	661	366	294

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

Robust standard errors accounting for clustering in parentheses.

Table B.7: *Estimated Impacts of Large School Size under Different Computation Strategies, Grade 4*

		Set 1 of Covariates			Set 2 of Covariates		
		Naïve difference	Ordinary least squares	Ordinary least squares on the common support	Naïve difference	Ordinary least squares	Ordinary least squares on the common support
French 2010	Estimated Effect	0.180**	-0.062	-0.064	0.180**	-0.032	-0.034
		(0.086)	(0.065)	(0.065)	(0.086)	(0.074)	(0.075)
	Observations	2,192	2,192	1,813	2,192	717	622
Mathematics 2010	Estimated Effect	0.210**	-0.054	-0.056	0.210**	-0.099	-0.114
		(0.088)	(0.067)	(0.066)	(0.088)	(0.078)	(0.079)
	Observations	2,192	2,192	1,813	2,192	717	622
Oral 2010	Estimated Effect	0.158*	-0.076	-0.096	0.158*	-0.120	-0.140*
		(0.090)	(0.078)	(0.077)	(0.090)	(0.084)	(0.084)
	Observations	1,102	1,102	906	1,102	638	554
Index 2010	Estimated Effect	0.223**	-0.075	-0.088	0.223**	-0.097	-0.108
		(0.095)	(0.070)	(0.069)	(0.095)	(0.073)	(0.074)
	Observations	1,097	1,097	901	1,097	633	549
French 2011	Estimated Effect	0.058	-0.089	-0.093	0.058	-0.121	-0.119
		(0.098)	(0.091)	(0.088)	(0.098)	(0.105)	(0.104)
	Observations	1,334	1,334	1,119	1,334	453	398
Mathematics 2011	Estimated Effect	0.066	-0.045	-0.038	0.066	-0.116	-0.120
		(0.103)	(0.096)	(0.092)	(0.103)	(0.110)	(0.109)
	Observations	1,334	1,334	1,119	1,334	453	398
Oral 2011	Estimated Effect	-0.030	-0.213*	-0.240**	-0.030	-0.330***	-0.307**
		(0.115)	(0.112)	(0.111)	(0.115)	(0.124)	(0.124)
	Observations	685	685	571	685	398	346
Index 2011	Estimated Effect	0.022	-0.190*	-0.207**	0.022	-0.230**	-0.214*
		(0.128)	(0.104)	(0.102)	(0.128)	(0.110)	(0.109)
	Observations	655	655	545	655	378	329

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

Robust standard errors accounting for clustering in parentheses.

Table B.8: *Models of the Probability of Attending a School of Large Size (Propensity Score) using the Full Set of Covariates*

	Grade 2	Grade 4
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001*	0.001
	(0.000)	(0.000)
The school is located in an urban area	0.230	0.445***
	(0.211)	(0.211)
Distance of the school to the nearest urban centre (in kilometres)	-0.001	-0.006
	(0.003)	(0.003)
Number of inhabitants in the city/village of the school	0.000*	0.000*
	(0.000)	(0.000)
There is a health centre in the locality of the school	0.713***	0.702***
	(0.239)	(0.228)
Age of the head teacher (in years)	0.027	0.043
	(0.028)	(0.028)
The head teacher is a female	-0.128	-0.085
	(0.567)	(0.605)
The head teacher has a BFEM or less	-0.360*	-0.291
	(0.189)	(0.181)
Experience of the head teacher (in years)	0.045	0.038
	(0.027)	(0.028)
The school was selected in 2009 for a school grant project	0.112	0.067
	(0.167)	(0.165)
Age of the teacher (in years)	-0.010	-0.037***
	(0.012)	(0.013)
The teacher is a female	0.078	0.023
	(0.172)	(0.174)
The teacher has a Baccalaureate/BSEN degree	0.157	0.081
	(0.173)	(0.163)
The teaching language is French	-0.118	-0.134
	(0.233)	(0.264)
The student is a female	0.019	-0.037
	(0.084)	(0.078)
Age of the student (in years)	-0.020	-0.017
	(0.061)	(0.055)
Time the student spends daily on housework (in hours)	0.012	0.014
	(0.039)	(0.038)
Student's household size	0.035**	-0.019
	(0.015)	(0.018)
The student goes to school regularly	-0.231	-0.072
	(0.210)	(0.193)
The student has a chronic disability	-0.267	0.420*
	(0.241)	(0.228)
The student works after school to earn money	-0.125	0.090
	(0.371)	(0.406)
The student has help with school work at home	0.161	-0.041
	(0.176)	(0.207)
Distance from the student's home to school (in kilometres)	0.054	0.035
	(0.072)	(0.065)
The student attends the closest school to his/her home	-0.123	0.049
	(0.233)	(0.239)
The student walks to school	0.752**	0.063
	(0.380)	(0.357)
Travel time from student's home to school (in minutes)	-0.011**	-0.006
	(0.006)	(0.006)
The student lives with his/her biological parents	0.200	0.255*
	(0.137)	(0.142)
The head of household is a female	0.316	0.185
	(0.212)	(0.205)
Age of the head of household (in years)	0.000	-0.004
	(0.005)	(0.006)

Table B.8: *Continued*

	Grade 2	Grade 4
The head of household has no formal education	-0.066	0.074
	(0.235)	(0.260)
The head of household attended koranic school or has completed primary school	-0.165	0.266
	(0.206)	(0.204)
The head of household is literate	-0.169	-0.042
	(0.155)	(0.160)
Socioeconomic status of the student's household	0.117***	-0.004
	(0.037)	(0.034)
The head of household is jobless	-0.644**	-0.381
	(0.302)	(0.314)
The head of household works in the public sector	-0.572	0.320
	(0.350)	(0.355)
The head of household works in the formal sector or the informal private sector	-0.595**	0.051
	(0.273)	(0.270)
The head of household is monogamous	0.248*	-0.133
	(0.139)	(0.133)
Standardized value of the prior French test	0.019	-0.154
	(0.103)	(0.106)
Standardized value of the prior mathematics test	-0.104	0.089
	(0.100)	(0.104)
Intercept	-3.869***	-2.664**
	(1.346)	(1.355)
Observations	785	794
Wald Chi2(39)	161.632	146.750
Prob > Chi2	0.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.9: *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Full Set of Covariates, Grade 2*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001*	-0.000
	(0.000)	(0.001)
The school is located in an urban area	0.230	0.139
	(0.211)	(0.235)
Distance of the school to the nearest urban centre (in kilometres)	-0.001	0.000
	(0.003)	(0.005)
Number of inhabitants in the city/village of the school	0.000*	-0.000
	(0.000)	(0.000)
There is a health centre in the locality of the school	0.713***	0.435
	(0.239)	(0.391)
Age of the head teacher (in years)	0.027	-0.025
	(0.028)	(0.035)
The head teacher is a female	-0.128	-0.020
	(0.567)	(0.562)
The head teacher has a BFEM or less	-0.360*	0.067
	(0.189)	(0.224)
Experience of the head teacher (in years)	0.045	0.021
	(0.027)	(0.033)
The school was selected in 2009 for a school grant project	0.112	-0.049
	(0.167)	(0.190)
Age of the teacher (in years)	-0.010	-0.008
	(0.012)	(0.014)
The teacher is a female	0.078	-0.018
	(0.172)	(0.206)
The teacher has a Baccalaureate/BSEN degree	0.157	-0.050
	(0.173)	(0.199)
The teaching language is French	-0.118	0.092
	(0.233)	(0.268)
The student is a female	0.019	-0.019
	(0.084)	(0.103)
Age of the student (in years)	-0.020	0.032
	(0.061)	(0.073)
Time the student spends daily on housework (in hours)	0.012	-0.019
	(0.039)	(0.050)
Student's household size	0.035**	0.009
	(0.015)	(0.019)
The student goes to school regularly	-0.231	0.102
	(0.210)	(0.263)
The student has a chronic disability	-0.267	-0.173
	(0.241)	(0.283)
The student works after school to earn money	-0.125	0.168
	(0.371)	(0.665)
The student has help with school work at home	0.161	-0.220
	(0.176)	(0.205)
Distance from the student's home to school (in kilometres)	0.054	0.027
	(0.072)	(0.094)
The student attends the closest school to his/her home	-0.123	-0.031
	(0.233)	(0.252)
The student walks to school	0.752**	-0.325
	(0.380)	(0.395)
Travel time from student's home to school (in minutes)	-0.011**	-0.000
	(0.006)	(0.006)
The student lives with his/her biological parents	0.200	0.068
	(0.137)	(0.171)
The head of household is a female	0.316	0.129
	(0.212)	(0.250)
Age of the head of household (in years)	0.000	0.001
	(0.005)	(0.006)

Table B.9: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	-0.066	-0.022
	(0.235)	(0.264)
The head of household attended koranic school or has completed primary school	-0.165	-0.012
	(0.206)	(0.237)
The head of household is literate	-0.169	0.106
	(0.155)	(0.189)
Socioeconomic status of the student's household	0.117***	-0.025
	(0.037)	(0.044)
The head of household is jobless	-0.644*	0.213
	(0.302)	(0.391)
The head of household works in the public sector	-0.572	0.092
	(0.350)	(0.432)
The head of household works in the formal sector or the informal private sector	-0.595**	0.083
	(0.273)	(0.355)
The head of household is monogamous	0.248*	0.018
	(0.139)	(0.164)
Standardized value of the prior French test	0.019	-0.050
	(0.103)	(0.112)
Standardized value of the prior mathematics test	-0.104	0.108
	(0.100)	(0.106)
Intercept	-3.869***	0.232
	(1.346)	(1.742)
Observations	785	461
p-value of the test of joint significance of differences	0.000	1.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.10: *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Full Set of Covariates, Grade 4*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001	0.000
	(0.000)	(0.001)
The school is located in an urban area	0.445**	0.249
	(0.211)	(0.253)
Distance of the school to the nearest urban centre (in kilometres)	-0.006	0.005
	(0.003)	(0.006)
Number of inhabitants in the city/village of the school	0.000*	0.000
	(0.000)	(0.000)
There is a health centre in the locality of the school	0.702***	-0.290
	(0.228)	(0.406)
Age of the head teacher (in years)	0.043	-0.032
	(0.028)	(0.039)
The head teacher is a female	-0.085	-0.479
	(0.605)	(0.546)
The head teacher has a BFEM or less	-0.291	0.095
	(0.181)	(0.238)
Experience of the head teacher (in years)	0.038	0.011
	(0.028)	(0.036)
The school was selected in 2009 for a school grant project	0.067	0.177
	(0.165)	(0.206)
Age of the teacher (in years)	-0.037***	0.004
	(0.013)	(0.015)
The teacher is a female	0.023	-0.029
	(0.174)	(0.204)
The teacher has a Baccalaureate/BSEN degree	0.081	-0.018
	(0.163)	(0.211)
The teaching language is French	-0.134	0.163
	(0.264)	(0.361)
The student is a female	-0.037	-0.117
	(0.078)	(0.104)
Age of the student (in years)	-0.017	-0.018
	(0.055)	(0.070)
Time the student spends daily on housework (in hours)	0.014	0.001
	(0.038)	(0.051)
Student's household size	-0.019	-0.020
	(0.018)	(0.023)
The student goes to school regularly	-0.072	-0.008
	(0.193)	(0.254)
The student has a chronic disability	0.420*	-0.111
	(0.228)	(0.269)
The student works after school to earn money	0.090	0.523
	(0.406)	(0.646)
The student has help with school work at home	-0.041	-0.075
	(0.207)	(0.230)
Distance from the student's home to school (in kilometres)	0.035	-0.039
	(0.065)	(0.079)
The student attends the closest school to his/her home	0.049	-0.031
	(0.239)	(0.264)
The student walks to school	0.063	-0.059
	(0.357)	(0.429)
Travel time from student's home to school (in minutes)	-0.006	-0.003
	(0.006)	(0.008)
The student lives with his/her biological parents	0.255*	0.085
	(0.142)	(0.173)
The head of household is a female	0.185	-0.010
	(0.205)	(0.251)
Age of the head of household (in years)	-0.004	-0.003
	(0.006)	(0.007)

Table B.10: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	0.074	0.159
	(0.260)	(0.287)
The head of household attended koranic school or has completed primary school	0.266	0.155
	(0.204)	(0.232)
The head of household is literate	-0.042	0.156
	(0.160)	(0.208)
Socioeconomic status of the student's household	-0.004	-0.003
	(0.034)	(0.038)
The head of household is jobless	-0.381	0.003
	(0.314)	(0.385)
The head of household works in the public sector	0.320	0.243
	(0.355)	(0.379)
The head of household works in the formal sector or the informal private sector	0.051	-0.036
	(0.270)	(0.324)
The head of household is monogamous	-0.133	0.004
	(0.133)	(0.159)
Standardized value of the prior French test	-0.154	-0.093
	(0.106)	(0.128)
Standardized value of the prior mathematics test	0.089	0.068
	(0.104)	(0.125)
Intercept	-2.664**	-0.008
	(1.355)	(1.676)
Observations	794	476
p-value of the test of joint significance of differences	0.000	1.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.11: *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Full Set of Covariates, Grade 2*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001 (0.000)	0.000 (0.001)
The school is located in an urban area	0.230 (0.211)	0.101 (0.239)
Distance of the school to the nearest urban centre (in kilometres)	-0.001 (0.003)	0.003 (0.005)
Number of inhabitants in the city/village of the school	0.000* (0.000)	0.000 (0.000)
There is a health centre in the locality of the school	0.713*** (0.239)	0.175 (0.456)
Age of the head teacher (in years)	0.027 (0.028)	-0.007 (0.038)
The head teacher is a female	-0.128 (0.567)	-0.251 (0.521)
The head teacher has a BFEM or less	-0.360* (0.189)	0.052 (0.231)
Experience of the head teacher (in years)	0.045 (0.027)	-0.005 (0.035)
The school was selected in 2009 for a school grant project	0.112 (0.167)	0.179 (0.200)
Age of the teacher (in years)	-0.010 (0.012)	-0.000 (0.014)
The teacher is a female	0.078 (0.172)	-0.065 (0.217)
The teacher has a Baccalaureate/BSEN degree	0.157 (0.173)	-0.107 (0.209)
The teaching language is French	-0.118 (0.233)	0.237 (0.298)
The student is a female	0.019 (0.084)	0.056 (0.101)
Age of the student (in years)	-0.020 (0.061)	0.084 (0.075)
Time the student spends daily on housework (in hours)	0.012 (0.039)	-0.000 (0.056)
Student's household size	0.035** (0.015)	0.030 (0.020)
The student goes to school regularly	-0.231 (0.210)	0.139 (0.268)
The student has a chronic disability	-0.267 (0.241)	-0.169 (0.290)
The student works after school to earn money	-0.125 (0.371)	1.127* (0.658)
The student has help with school work at home	0.161 (0.176)	-0.041 (0.213)
Distance from the student's home to school (in kilometres)	0.054 (0.072)	0.056 (0.096)
The student attends the closest school to his/her home	-0.123 (0.233)	-0.198 (0.248)
The student walks to school	0.752** (0.380)	0.080 (0.394)
Travel time from student's home to school (in minutes)	-0.011** (0.006)	-0.008 (0.008)
The student lives with his/her biological parents	0.200 (0.137)	0.046 (0.184)
The head of household is a female	0.316 (0.212)	0.283 (0.253)
Age of the head of household (in years)	0.000 (0.005)	-0.003 (0.006)

Table B.11: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	-0.066	-0.101
	(0.235)	(0.283)
The head of household attended koranic school or has completed primary school	-0.165	0.033
	(0.206)	(0.252)
The head of household is literate	-0.169	-0.199
	(0.155)	(0.194)
Socioeconomic status of the student's household	0.117***	-0.001
	(0.037)	(0.048)
The head of household is jobless	-0.644**	0.020
	(0.302)	(0.397)
The head of household works in the public sector	-0.572	0.077
	(0.350)	(0.445)
The head of household works in the formal sector or the informal private sector	-0.595**	0.026
	(0.273)	(0.361)
The head of household is monogamous	0.248*	0.158
	(0.139)	(0.171)
Standardized value of the prior French test	0.019	-0.026
	(0.103)	(0.115)
Standardized value of the prior mathematics test	-0.104	0.047
	(0.100)	(0.110)
Intercept	-3.869***	-0.927
	(1.346)	(1.885)
Observations	785	461
p-value of the test of joint significance of differences	0.000	0.999

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.12: *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Full Set of Covariates, Grade 4*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001 (0.000)	0.000 (0.001)
The school is located in an urban area	0.445** (0.211)	0.249 (0.253)
Distance of the school to the nearest urban centre (in kilometres)	-0.006 (0.003)	0.005 (0.006)
Number of inhabitants in the city/village of the school	0.000* (0.000)	0.000 (0.000)
There is a health centre in the locality of the school	0.702*** (0.228)	-0.290 (0.406)
Age of the head teacher (in years)	0.043 (0.028)	-0.032 (0.039)
The head teacher is a female	-0.085 (0.605)	-0.479 (0.546)
The head teacher has a BFEM or less	-0.291 (0.181)	0.095 (0.238)
Experience of the head teacher (in years)	0.038 (0.028)	0.011 (0.036)
The school was selected in 2009 for a school grant project	0.067 (0.165)	0.177 (0.206)
Age of the teacher (in years)	-0.037*** (0.013)	0.004 (0.015)
The teacher is a female	0.023 (0.174)	-0.029 (0.204)
The teacher has a Baccalaureate/BSEN degree	0.081 (0.163)	-0.018 (0.211)
The teaching language is French	-0.134 (0.264)	0.163 (0.361)
The student is a female	-0.037 (0.078)	-0.117 (0.104)
Age of the student (in years)	-0.017 (0.055)	-0.018 (0.070)
Time the student spends daily on housework (in hours)	0.014 (0.038)	0.001 (0.051)
Student's household size	-0.019 (0.018)	-0.020 (0.023)
The student goes to school regularly	-0.072 (0.193)	-0.008 (0.254)
The student has a chronic disability	0.420* (0.228)	-0.111 (0.269)
The student works after school to earn money	0.090 (0.406)	0.523 (0.646)
The student has help with school work at home	-0.041 (0.207)	-0.075 (0.230)
Distance from the student's home to school (in kilometres)	0.035 (0.065)	-0.039 (0.079)
The student attends the closest school to his/her home	0.049 (0.239)	-0.031 (0.264)
The student walks to school	0.063 (0.357)	-0.059 (0.429)
Travel time from student's home to school (in minutes)	-0.006 (0.006)	-0.003 (0.008)
The student lives with his/her biological parents	0.255* (0.142)	0.085 (0.173)
The head of household is a female	0.185 (0.205)	-0.010 (0.251)
Age of the head of household (in years)	-0.004 (0.006)	-0.003 (0.007)

Table B.12: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	0.074	-0.072
	(0.260)	(0.292)
The head of household attended koranic school or has completed primary school	0.266	0.071
	(0.204)	(0.241)
The head of household is literate	-0.042	-0.015
	(0.160)	(0.217)
Socioeconomic status of the student's household	-0.004	-0.020
	(0.034)	(0.042)
The head of household is jobless	-0.381	-0.039
	(0.314)	(0.412)
The head of household works in the public sector	0.320	0.006
	(0.355)	(0.413)
The head of household works in the formal sector or the informal private sector	0.051	-0.064
	(0.270)	(0.363)
The head of household is monogamous	-0.133	0.029
	(0.133)	(0.162)
Standardized value of the prior French test	-0.154	-0.141
	(0.106)	(0.138)
Standardized value of the prior mathematics test	0.089	0.089
	(0.104)	(0.130)
Intercept	-2.664**	1.848
	(1.355)	(1.881)
Observations	794	476
p-value of the test of joint significance of differences	0.000	0.999

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.13: *Independent Balance Tests: T-statistics from Tests of Equality of Unweighted Means and Weighted Means of Covariates using a Propensity Score Predicted with the Full Set of Covariates*

	Grade 2			Grade 4		
	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET
Distance from the school to the country's capital (Dakar) (in kilometres)	-2.07	0.12	1.28	-2.66	-0.09	0.51
The school is located in an urban area	7.47	0.09	-0.70	7.68	0.17	0.46
Distance of the school to the nearest urban centre (in kilometres)	-4.29	0.27	1.04	-4.32	0.04	0.65
Number of inhabitants in the city/village of the school	4.23	-0.30	-0.67	4.19	0.62	0.57
There is a health centre in the locality of the school	8.68	1.14	-0.12	8.90	-0.34	-0.31
Age of the head teacher (in years)	11.78	-0.33	-0.86	12.51	0.12	-0.90
The head teacher is a female	1.42	-0.19	-0.70	1.57	-0.24	-0.33
The head teacher has a BFEM or less	-1.84	0.24	0.10	-2.03	0.71	-0.07
Experience of the head teacher (in years)	12.47	0.02	-0.95	13.39	0.31	-0.71
The school was selected in 2009 for a school grant project	1.00	-0.11	0.99	0.78	-0.55	1.09
Age of the teacher (in years)	2.34	-0.50	0.09	1.46	0.31	0.22
The teacher is a female	3.29	0.09	-0.46	2.09	0.54	0.18
The teacher has a Baccalaureate/BSEN degree	-0.59	-0.29	-0.38	1.41	0.30	-0.81
The teaching language is French	0.87	0.18	0.58	0.28	0.11	0.26
The student is a female	0.29	-0.36	-0.06	-0.57	-0.40	-0.89
Age of the student (in years)	-0.14	0.53	1.04	0.47	0.06	-0.03
Time the student spends daily on housework (in hours)	-2.88	-0.05	0.69	-1.33	0.05	-0.22
Student's household size	-0.80	0.36	1.42	-3.02	-0.67	-0.80
The student goes to school regularly	-1.49	-0.00	0.65	-0.91	0.13	0.25
The student has a chronic disability	-1.03	-0.21	0.09	1.29	0.05	-0.37
The student works after school to earn money	-0.64	0.22	1.00	-0.42	0.22	0.61
The student has help with school work at home	3.75	-0.99	-0.54	3.39	0.05	-0.21
Distance from the student's home to school (in kilometres)	-2.26	0.26	-0.12	1.15	-0.09	-0.30
The student attends the closest school to his/her home	-3.02	0.00	-0.03	-3.13	-0.22	-0.10
The student walks to school	1.54	-0.66	0.30	-0.85	-0.02	0.23
Travel time from student's home to school (in minutes)	-0.94	0.06	-0.89	-0.31	-0.02	-0.19
The student lives with his/her biological parents	0.47	0.22	-0.19	1.20	-0.21	0.84
The head of household is a female	2.02	0.26	0.88	2.49	0.15	-0.43
Age of the head of household (in years)	0.26	0.01	-0.65	-0.68	-0.14	-0.20

Table B.13: *Continued*

	Grade 2			Grade 4		
	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET
The head of household has no formal education	-1.05	-0.04	0.53	-2.41	-0.28	-0.38
The head of household attended koranic school or has completed primary school	-2.59	0.17	0.35	0.26	0.27	0.68
The head of household has at least primary education level	4.53	-0.16	-0.82	2.70	0.04	-0.23
The head of household is literate	1.55	0.30	-1.55	2.85	0.66	0.19
Socioeconomic status of the student's household	8.62	-0.69	-0.76	6.62	0.23	-0.39
The head of household is jobless	1.18	0.42	-0.29	0.42	-0.09	-0.26
The head of household works in the public sector	3.31	-0.03	-0.54	2.86	0.93	0.59
The head of household works in the formal sector or the informal private sector	-4.62	-0.02	0.86	-2.81	-0.64	-0.32
The head of household is retired	2.24	-0.54	-0.24	0.52	0.11	0.22
The head of household is monogamous	1.99	-0.09	0.12	0.47	0.13	0.37
Standardized value of the prior French test	3.17	0.05	-0.55	1.92	-0.15	-0.58
Standardized value of the prior mathematics test	2.94	0.66	-0.41	1.24	0.09	-0.33

Table B.14: *Independent Balance Tests: Normalized Differences between Small and Large Schools*

	Grade 2	Grade 4
Distance from the school to the country's capital (Dakar) (in kilometres)	-0.23	-0.30
The school is located in an urban area	0.88	0.91
Distance of the school to the nearest urban centre (in kilometres)	-0.44	-0.45
Number of inhabitants in the city/village of the school	0.53	0.54
There is a health centre in the locality of the school	0.83	0.83
Age of the head teacher (in years)	1.21	1.27
The head teacher is a female	0.17	0.19
The head teacher has a BFEM or less	-0.20	-0.22
Experience of the head teacher (in years)	1.30	1.37
The school was selected in 2009 for a school grant project	0.11	0.09
Age of the teacher (in years)	0.27	0.16
The teacher is a female	0.36	0.24
The teacher has a Baccalaureate/BSEN degree	-0.07	0.16
The teaching language is French	0.10	0.03
The student is a female	0.01	-0.01
Age of the student (in years)	-0.01	0.03
Time the student spends daily on housework (in hours)	-0.23	-0.12
Student's household size	-0.07	-0.26
The student goes to school regularly	-0.15	-0.08
The student has a chronic disability	-0.08	0.10
The student works after school to earn money	-0.05	-0.03
The student has help with school work at home	0.36	0.35
Distance from the student's home to school (in kilometres)	-0.18	0.10
The student attends the closest school to his/her home	-0.29	-0.27
The student walks to school	0.11	-0.07
Travel time from student's home to school (in minutes)	-0.08	-0.03
The student lives with his/her biological parents	0.04	0.09
The head of household is a female	0.19	0.21
Age of the head of household (in years)	0.02	-0.06
The head of household has no formal education	-0.09	-0.21
The head of household attended koranic school or has completed primary school	-0.20	0.02
The head of household has at least primary education level	0.37	0.25
The head of household is literate	0.13	0.25
Socioeconomic status of the student's household	0.83	0.66
The head of household is jobless	0.09	0.04
The head of household works in the public sector	0.30	0.27
The head of household works in the formal sector or the informal private sector	-0.39	-0.23
The head of household is retired	0.21	0.04
The head of household is monogamous	0.16	0.04
Standardized value of the prior French test	0.27	0.17
Standardized value of the prior mathematics test	0.24	0.11

Table B.15: *Models of the Probability of Attending a School of Large Size (Propensity Score) using the Reduced Set of Covariates*

	Grade 2	Grade 4
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001*	0.000
	(0.000)	(0.000)
The school is located in an urban area	0.311	0.356*
	(0.194)	(0.201)
Distance of the school to the nearest urban centre (in kilometres)	-0.002	-0.004
	(0.003)	(0.003)
Number of inhabitants in the city/village of the school	0.000	0.000*
	(0.000)	(0.000)
There is a health centre in the locality of the school	0.703***	0.705***
	(0.231)	(0.223)
Age of the head teacher (in years)	0.018	0.029
	(0.028)	(0.028)
The head teacher is a female	-0.170	0.101
	(0.574)	(0.576)
The head teacher has a BFEM or less	-0.368**	-0.321*
	(0.182)	(0.178)
Experience of the head teacher (in years)	0.047*	0.046*
	(0.027)	(0.028)
The school was selected in 2009 for a school grant project	0.107	0.096
	(0.164)	(0.161)
Age of the teacher (in years)	-0.004	-0.032**
	(0.012)	(0.013)
The teacher is a female	0.130	0.104
	(0.161)	(0.172)
The teacher has a Baccalaureate/BSEN degree	0.085	0.113
	(0.167)	(0.162)
The teaching language is French	-0.075	-0.012
	(0.220)	(0.273)
The student is a female	0.013	0.012
	(0.034)	(0.037)
Age of the student (in years)	0.032	-0.006
	(0.048)	(0.043)
Standardized value of the prior French test	0.085	0.010
	(0.083)	(0.080)
Standardized value of the prior mathematics test	-0.057	-0.048
	(0.071)	(0.083)
Intercept	-3.457***	-2.694***
	(0.984)	(0.999)
Observations	2,284	2,269
Wald Chi2(39)	107.777	133.436
Prob > Chi2	0.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.16: *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 2*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001*	0.001
	(0.000)	(0.001)
The school is located in an urban area	0.230	-0.016
	(0.211)	(0.225)
Distance of the school to the nearest urban centre (in kilometres)	-0.001	0.002
	(0.003)	(0.006)
Number of inhabitants in the city/village of the school	0.000*	0.000
	(0.000)	(0.000)
There is a health centre in the locality of the school	0.713***	0.181
	(0.239)	(0.423)
Age of the head teacher (in years)	0.027	0.004
	(0.028)	(0.035)
The head teacher is a female	-0.128	0.057
	(0.567)	(0.527)
The head teacher has a BFEM or less	-0.360*	0.056
	(0.189)	(0.219)
Experience of the head teacher (in years)	0.045	0.003
	(0.027)	(0.032)
The school was selected in 2009 for a school grant project	0.112	-0.159
	(0.167)	(0.186)
Age of the teacher (in years)	-0.010	-0.011
	(0.012)	(0.013)
The teacher is a female	0.078	-0.104
	(0.172)	(0.201)
The teacher has a Baccalaureate/BSEN degree	0.157	0.110
	(0.173)	(0.192)
The teaching language is French	-0.118	-0.159
	(0.233)	(0.285)
The student is a female	0.019	0.042
	(0.084)	(0.101)
Age of the student (in years)	-0.020	-0.055
	(0.061)	(0.071)
Time the student spends daily on housework (in hours)	0.012	-0.011
	(0.039)	(0.048)
Student's household size	0.035**	0.048***
	(0.015)	(0.018)
The student goes to school regularly	-0.231	-0.311
	(0.210)	(0.244)
The student has a chronic disability	-0.267	-0.217
	(0.241)	(0.247)
The student works after school to earn money	-0.125	0.122
	(0.371)	(0.651)
The student has help with school work at home	0.161	0.090
	(0.176)	(0.196)
Distance from the student's home to school (in kilometres)	0.054	0.081
	(0.072)	(0.089)
The student attends the closest school to his/her home	-0.123	-0.054
	(0.233)	(0.227)
The student walks to school	0.752**	0.813**
	(0.380)	(0.398)
Travel time from student's home to school (in minutes)	-0.011**	-0.015**
	(0.006)	(0.006)
The student lives with his/her biological parents	0.200	0.239
	(0.137)	(0.156)
The head of household is a female	0.316	0.434*
	(0.212)	(0.247)
Age of the head of household (in years)	0.000	-0.001
	(0.005)	(0.005)

Table B.16: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	-0.066	-0.019
	(0.235)	(0.250)
The head of household attended koranic school or has completed primary school	-0.165	-0.194
	(0.206)	(0.223)
The head of household is literate	-0.169	-0.111
	(0.155)	(0.176)
Socioeconomic status of the student's household	0.117***	0.116***
	(0.037)	(0.043)
The head of household is jobless	-0.644*	-0.824**
	(0.302)	(0.368)
The head of household works in the public sector	-0.572	-0.723*
	(0.350)	(0.416)
The head of household works in the formal sector or the informal private sector	-0.595**	-0.789**
	(0.273)	(0.343)
The head of household is monogamous	0.248*	0.278*
	(0.139)	(0.155)
Standardized value of the prior French test	0.019	-0.082
	(0.103)	(0.108)
Standardized value of the prior mathematics test	-0.104	0.021
	(0.100)	(0.104)
Intercept	-3.869***	-0.235
	(1.346)	(1.778)
Observations	785	511
p-value of the test of joint significance of differences	0.000	0.034

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.17: *Joint Balance Test for the Estimation of the ATE using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 4*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001 (0.000)	0.000 (0.001)
The school is located in an urban area	0.445** (0.211)	0.116 (0.231)
Distance of the school to the nearest urban centre (in kilometres)	-0.006 (0.003)	0.001 (0.006)
Number of inhabitants in the city/village of the school	0.000* (0.000)	-0.000 (0.000)
There is a health centre in the locality of the school	0.702*** (0.228)	-0.084 (0.417)
Age of the head teacher (in years)	0.043 (0.028)	-0.002 (0.036)
The head teacher is a female	-0.085 (0.605)	-0.104 (0.579)
The head teacher has a BFEM or less	-0.291 (0.181)	0.182 (0.220)
Experience of the head teacher (in years)	0.038 (0.028)	-0.001 (0.033)
The school was selected in 2009 for a school grant project	0.067 (0.165)	-0.161 (0.190)
Age of the teacher (in years)	-0.037*** (0.013)	0.005 (0.014)
The teacher is a female	0.023 (0.174)	-0.118 (0.195)
The teacher has a Baccalaureate/BSEN degree	0.081 (0.163)	0.052 (0.192)
The teaching language is French	-0.134 (0.264)	-0.125 (0.311)
The student is a female	-0.037 (0.078)	-0.101 (0.088)
Age of the student (in years)	-0.017 (0.055)	-0.009 (0.062)
Time the student spends daily on housework (in hours)	0.014 (0.038)	0.015 (0.046)
Student's household size	-0.019 (0.018)	-0.019 (0.021)
The student goes to school regularly	-0.072 (0.193)	0.020 (0.240)
The student has a chronic disability	0.420* (0.228)	0.428* (0.254)
The student works after school to earn money	0.090 (0.406)	0.394 (0.583)
The student has help with school work at home	-0.041 (0.207)	0.051 (0.219)
Distance from the student's home to school (in kilometres)	0.035 (0.065)	0.045 (0.079)
The student attends the closest school to his/her home	0.049 (0.239)	0.047 (0.247)
The student walks to school	0.063 (0.357)	0.204 (0.412)
Travel time from student's home to school (in minutes)	-0.006 (0.006)	-0.008 (0.007)
The student lives with his/her biological parents	0.255* (0.142)	0.245 (0.156)
The head of household is a female	0.185 (0.205)	0.094 (0.239)
Age of the head of household (in years)	-0.004 (0.006)	-0.006 (0.007)

Table B.17: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	0.074	0.216
	(0.260)	(0.274)
The head of household attended koranic school or has completed primary school	0.266	0.347
	(0.204)	(0.227)
The head of household is literate	-0.042	0.014
	(0.160)	(0.198)
Socioeconomic status of the student's household	-0.004	0.008
	(0.034)	(0.037)
The head of household is jobless	-0.381	-0.484
	(0.314)	(0.358)
The head of household works in the public sector	0.320	0.247
	(0.355)	(0.383)
The head of household works in the formal sector or the informal private sector	0.051	-0.044
	(0.270)	(0.320)
The head of household is monogamous	-0.133	-0.042
	(0.133)	(0.157)
Standardized value of the prior French test	-0.154	-0.195
	(0.106)	(0.125)
Standardized value of the prior mathematics test	0.089	0.156
	(0.104)	(0.123)
Intercept	-2.664**	0.172
	(1.355)	(1.672)
Observations	794	480
p-value of the test of joint significance of differences	0.000	0.664

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.18: *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 2*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001*	0.001
	(0.000)	(0.001)
The school is located in an urban area	0.230	0.063
	(0.211)	(0.232)
Distance of the school to the nearest urban centre (in kilometres)	-0.001	0.006
	(0.003)	(0.006)
Number of inhabitants in the city/village of the school	0.000*	0.000
	(0.000)	(0.000)
There is a health centre in the locality of the school	0.713***	-0.101
	(0.239)	(0.485)
Age of the head teacher (in years)	0.027	0.006
	(0.028)	(0.037)
The head teacher is a female	-0.128	-0.362
	(0.567)	(0.526)
The head teacher has a BFEM or less	-0.360*	0.085
	(0.189)	(0.232)
Experience of the head teacher (in years)	0.045	-0.023
	(0.027)	(0.035)
The school was selected in 2009 for a school grant project	0.112	0.043
	(0.167)	(0.194)
Age of the teacher (in years)	-0.010	-0.000
	(0.012)	(0.014)
The teacher is a female	0.078	-0.091
	(0.172)	(0.207)
The teacher has a Baccalaureate/BSEN degree	0.157	0.046
	(0.173)	(0.201)
The teaching language is French	-0.118	0.035
	(0.233)	(0.304)
The student is a female	0.019	0.040
	(0.084)	(0.099)
Age of the student (in years)	-0.020	0.001
	(0.061)	(0.072)
Time the student spends daily on housework (in hours)	0.012	0.033
	(0.039)	(0.054)
Student's household size	0.035**	0.059***
	(0.015)	(0.020)
The student goes to school regularly	-0.231	-0.260
	(0.210)	(0.235)
The student has a chronic disability	-0.267	-0.355
	(0.241)	(0.261)
The student works after school to earn money	-0.125	1.139*
	(0.371)	(0.674)
The student has help with school work at home	0.161	0.186
	(0.176)	(0.201)
Distance from the student's home to school (in kilometres)	0.054	0.091
	(0.072)	(0.089)
The student attends the closest school to his/her home	-0.123	-0.009
	(0.233)	(0.234)
The student walks to school	0.752**	0.642*
	(0.380)	(0.389)
Travel time from student's home to school (in minutes)	-0.011**	-0.014*
	(0.006)	(0.007)
The student lives with his/her biological parents	0.200	0.105
	(0.137)	(0.173)
The head of household is a female	0.316	0.395
	(0.212)	(0.258)
Age of the head of household (in years)	0.000	-0.004
	(0.005)	(0.006)

Table B.18: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	-0.066	-0.173
	(0.235)	(0.264)
The head of household attended koranic school or has completed primary school	-0.165	-0.167
	(0.206)	(0.237)
The head of household is literate	-0.169	-0.403**
	(0.155)	(0.181)
Socioeconomic status of the student's household	0.117***	0.127***
	(0.037)	(0.047)
The head of household is jobless	-0.644**	-0.873**
	(0.302)	(0.397)
The head of household works in the public sector	-0.572	-0.667
	(0.350)	(0.449)
The head of household works in the formal sector or the informal private sector	-0.595**	-0.732*
	(0.273)	(0.376)
The head of household is monogamous	0.248*	0.398**
	(0.139)	(0.162)
Standardized value of the prior French test	0.019	-0.026
	(0.103)	(0.109)
Standardized value of the prior mathematics test	-0.104	-0.055
	(0.100)	(0.103)
Intercept	-3.869***	-0.285
	(1.346)	(1.823)
Observations	785	511
p-value of the test of joint significance of differences	0.000	0.001

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.19: *Joint Balance Test for the Estimation of the ATET using the Propensity Score Predicted with the Reduced Set of Covariates, Grade 4*

	Unweighted Sample of Students	Weighted Sample of Students
Distance from the school to the country's capital (Dakar) (in kilometres)	0.001 (0.000)	0.000 (0.001)
The school is located in an urban area	0.445** (0.211)	0.273 (0.252)
Distance of the school to the nearest urban centre (in kilometres)	-0.006 (0.003)	0.003 (0.006)
Number of inhabitants in the city/village of the school	0.000* (0.000)	0.000 (0.000)
There is a health centre in the locality of the school	0.702*** (0.228)	-0.309 (0.446)
Age of the head teacher (in years)	0.043 (0.028)	-0.007 (0.038)
The head teacher is a female	-0.085 (0.605)	-0.609 (0.554)
The head teacher has a BFEM or less	-0.291 (0.181)	0.162 (0.241)
Experience of the head teacher (in years)	0.038 (0.028)	-0.016 (0.036)
The school was selected in 2009 for a school grant project	0.067 (0.165)	0.203 (0.204)
Age of the teacher (in years)	-0.037*** (0.013)	0.002 (0.014)
The teacher is a female	0.023 (0.174)	-0.159 (0.203)
The teacher has a Baccalaureate/BSEN degree	0.081 (0.163)	0.039 (0.208)
The teaching language is French	-0.134 (0.264)	0.034 (0.337)
The student is a female	-0.037 (0.078)	-0.107 (0.092)
Age of the student (in years)	-0.017 (0.055)	-0.054 (0.068)
Time the student spends daily on housework (in hours)	0.014 (0.038)	0.025 (0.049)
Student's household size	-0.019 (0.018)	-0.029 (0.021)
The student goes to school regularly	-0.072 (0.193)	-0.242 (0.252)
The student has a chronic disability	0.420* (0.228)	0.258 (0.249)
The student works after school to earn money	0.090 (0.406)	0.473 (0.596)
The student has help with school work at home	-0.041 (0.207)	-0.128 (0.217)
Distance from the student's home to school (in kilometres)	0.035 (0.065)	0.037 (0.077)
The student attends the closest school to his/her home	0.049 (0.239)	0.131 (0.261)
The student walks to school	0.063 (0.357)	0.023 (0.400)
Travel time from student's home to school (in minutes)	-0.006 (0.006)	-0.007 (0.007)
The student lives with his/her biological parents	0.255* (0.142)	0.297* (0.153)
The head of household is a female	0.185 (0.205)	-0.079 (0.246)
Age of the head of household (in years)	-0.004 (0.006)	-0.009 (0.007)

Table B.19: *Continued*

	Unweighted Sample of Students	Weighted Sample of Students
The head of household has no formal education	0.074	0.042
	(0.260)	(0.288)
The head of household attended koranic school or has completed primary school	0.266	0.254
	(0.204)	(0.240)
The head of household is literate	-0.042	-0.095
	(0.160)	(0.215)
Socioeconomic status of the student's household	-0.004	0.014
	(0.034)	(0.040)
The head of household is jobless	-0.381	-0.485
	(0.314)	(0.394)
The head of household works in the public sector	0.320	-0.090
	(0.355)	(0.435)
The head of household works in the formal sector or the informal private sector	0.051	-0.213
	(0.270)	(0.372)
The head of household is monogamous	-0.133	-0.026
	(0.133)	(0.162)
Standardized value of the prior French test	-0.154	-0.163
	(0.106)	(0.133)
Standardized value of the prior mathematics test	0.089	0.094
	(0.104)	(0.126)
Intercept	-2.664**	2.199
	(1.355)	(1.874)
Observations	794	480
p-value of the test of joint significance of differences	0.000	0.345

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.20: *Independent Balance Tests: T-statistics from Tests of Equality of Unweighted Means and Weighted Means of Covariates Using a Propensity Score Predicted with the Reduced Set of Covariates*

	Grade 2			Grade 4		
	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET
Distance from the school to the country's capital (Dakar) (in kilometres)	-2.07	0.72	1.36	-2.66	0.47	0.92
The school is located in an urban area	7.47	-0.28	-0.57	7.68	-0.19	-0.29
Distance of the school to the nearest urban centre (in kilometres)	-4.29	0.23	1.15	-4.32	0.39	1.10
Number of inhabitants in the city/village of the school	4.23	-0.22	-0.44	4.19	-0.29	-0.40
There is a health centre in the locality of the school	8.68	0.36	-0.28	8.90	-0.34	-0.42
Age of the head teacher (in years)	11.78	-0.76	-1.44	12.51	-0.74	-1.04
The head teacher is a female	1.42	0.22	-0.51	1.57	0.07	-0.42
The head teacher has a BFEM or less	-1.84	-0.00	-0.40	-2.03	0.33	0.17
Experience of the head teacher (in years)	12.47	-0.70	-1.50	13.39	-0.68	-1.11
The school was selected in 2009 for a school grant project	1.00	-1.03	-0.06	0.78	-0.42	1.37
Age of the teacher (in years)	2.34	-0.53	0.45	1.46	0.14	0.04
The teacher is a female	3.29	-0.04	-0.83	2.09	-0.36	-0.39
The teacher has a Baccalaureate/BSEN degree	-0.59	0.08	0.14	1.41	0.20	-0.20
The teaching language is French	0.87	-0.16	-0.08	0.28	0.10	0.32
The student is a female	0.29	0.07	-0.37	-0.57	-0.06	0.31
Age of the student (in years)	-0.14	0.11	0.20	0.47	-0.23	-0.17
Time the student spends daily on housework (in hours)	-2.88	-0.52	0.39	-1.33	0.44	0.77
Student's household size	-0.80	1.97	2.49	-3.02	-0.42	-0.82
The student goes to school regularly	-1.49	-0.92	-0.94	-0.91	0.53	-0.27
The student has a chronic disability	-1.03	-0.36	-1.06	1.29	1.41	0.81
The student works after school to earn money	-0.64	0.20	1.01	-0.42	0.17	0.24
The student has help with school work at home	3.75	0.75	0.81	3.39	-0.38	-1.00
Distance from the student's home to school (in kilometres)	-2.26	-1.16	-1.17	1.15	-0.36	0.02
The student attends the closest school to his/her home	-3.02	0.79	1.20	-3.13	0.76	0.99
The student walks to school	1.54	1.98	1.57	-0.85	0.58	0.33
Travel time from student's home to school (in minutes)	-0.94	-2.65	-2.33	-0.31	-1.18	-0.68
The student lives with his/her biological parents	0.47	1.40	0.34	1.20	2.11	-2.71
The head of household is a female	2.02	0.71	0.51	2.49	-0.60	-1.17
Age of the head of household (in years)	0.26	0.54	0.09	-0.68	-1.48	-1.18

Table B.20: *Continued*

	Grade 2			Grade 4		
	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET	Unweighted Sample	Weighted Sample of Students - ATE	Weighted Sample of Students - ATET
The head of household has no formal education	-1.05	0.11	-0.22	-2.41	-1.05	-1.03
The head of household attended koranic school or has completed primary school	-2.59	-1.62	-0.81	0.26	1.34	1.73
The head of household has at least primary education level	4.53	1.91	1.34	2.70	-0.20	-0.34
The head of household is literate	1.55	-0.43	-1.52	2.85	0.98	0.69
Socioeconomic status of the student's household	8.62	2.38	1.82	6.62	-0.13	-0.86
The head of household is jobless	1.18	-0.59	-0.93	0.42	-2.45	-1.94
The head of household works in the public sector	3.31	-1.24	0.95	2.86	1.58	1.19
The head of household works in the formal sector or the informal private sector	-4.62	-2.07	-1.30	-2.81	1.02	0.89
The head of household is retired	2.24	2.37	2.85	0.52	-0.51	0.27
The head of household is monogamous	1.99	0.85	1.15	0.47	-0.05	0.42
Standardized value of the prior French test	3.17	0.28	0.20	1.92	-0.48	-0.36
Standardized value of the prior mathematics test	2.94	0.33	-0.22	1.24	-0.45	-0.69

Table B.21: *Effect of Trimming on Sample Size for the Second Survey Written Test Scores*

		$\widehat{p}(X_i) < 0.10$		$0.10 \leqslant \widehat{p}(X_i) \leqslant 0.90$		$\widehat{p}(X_i) > 0.90$		Total
		Treatment	Control	Treatment	Control	Treatment	Control	
Grade 2	Whole sample (No household data)	22	778	579	830	0	0	2,209
	Reduced sample (With household data)	8	278	177	230	15	2	710
Grade 4	Whole sample (No household data)	16	843	569	752	6	6	2,192
	Reduced sample (With household data)	6	271	189	245	3	3	717

Table B.22: *Estimated ATE and ATET with the Method of Inverse Propensity Score Weighting by Grade*

		Grade 2				Grade 4			
		Set 1 of Covariates		Set 2 of Covariates		Set 1 of Covariates		Set 2 of Covariates	
		ATE	ATET	ATE	ATET	ATE	ATET	ATE	ATET
French 2010	Estimated effect	-0.020	-0.063	0.045	-0.162	-0.080	-0.267*	-0.034	-0.264*
		(0.112)	(0.113)	(0.144)	(0.149)	(0.114)	(0.161)	(0.140)	(0.159)
	Observations	1,409	1,409	407	424	1,321	1,333	434	440
Mathematics 2010	Estimated effect	-0.044	-0.135	0.019	-0.162	-0.093	-0.235	-0.105	-0.224
		(0.098)	(0.094)	(0.126)	(0.140)	(0.121)	(0.150)	(0.145)	(0.139)
	Observations	1,409	1,409	407	424	1,321	1,333	434	440
Oral 2010	Estimated effect	-0.062	-0.082	0.058	-0.110	-0.128	-0.218	-0.132	-0.353**
		(0.112)	(0.124)	(0.145)	(0.173)	(0.126)	(0.134)	(0.138)	(0.153)
	Observations	710	710	351	364	662	667	385	390
Index 2010	Estimated effect	-0.110	-0.176	0.023	-0.200	-0.097	-0.210	-0.118	-0.326**
		(0.120)	(0.124)	(0.140)	(0.163)	(0.132)	(0.144)	(0.152)	(0.153)
	Observations	702	702	344	357	659	664	382	387
French 2011	Estimated effect	-0.068	-0.189	0.044	-0.158	-0.132	-0.445**	-0.239	-0.487***
		(0.117)	(0.123)	(0.151)	(0.174)	(0.124)	(0.196)	(0.165)	(0.165)
	Observations	874	874	267	280	803	814	273	278
Mathematics 2011	Estimated effect	0.000	-0.132	0.063	-0.206	-0.083	-0.362*	-0.242	-0.437***
		(0.114)	(0.104)	(0.142)	(0.149)	(0.135)	(0.198)	(0.172)	(0.157)
	Observations	874	874	267	280	803	814	273	278
Oral 2011	Estimated effect	-0.021	-0.061	0.090	-0.175	-0.333**	-0.425***	-0.428**	-0.618***
		(0.143)	(0.167)	(0.178)	(0.227)	(0.153)	(0.146)	(0.191)	(0.161)
	Observations	435	435	213	221	418	423	236	240
Index 2011	Estimated effect	-0.073	-0.167	0.003	-0.233	-0.255	-0.457**	-0.375*	-0.584***
		(0.144)	(0.162)	(0.186)	(0.217)	(0.164)	(0.176)	(0.203)	(0.174)
	Observations	423	423	207	215	395	400	222	226

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.23: *Estimated ATE and ATET for Subpopulations, Grade 2*

		ATE				ATET			
		Boys	Girls	Rural zones	Urban zones	Boys	Girls	Rural zones	Urban zones
French 2010	Estimated effect	0.044	0.069	0.014	0.013	-0.066	0.036	-0.055	-0.060
		(0.129)	(0.117)	(0.136)	(0.123)	(0.118)	(0.117)	(0.140)	(0.124)
	Observations	183	211	206	145	191	221	209	175
Mathematics 2010	Estimated effect	0.092	0.064	0.128	0.001	0.061	-0.007	0.098	-0.064
		(0.111)	(0.100)	(0.097)	(0.095)	(0.108)	(0.084)	(0.104)	(0.107)
	Observations	183	211	206	145	191	221	209	175
Oral 2010	Estimated effect	0.127	0.007	0.158	0.260	0.091	-0.010	0.094	0.136
		(0.145)	(0.120)	(0.144)	(0.160)	(0.130)	(0.120)	(0.149)	(0.157)
	Observations	167	175	177	125	170	183	180	148
Index 2010	Estimated effect	0.048	0.077	0.124	0.033	-0.024	0.059	0.066	-0.046
		(0.134)	(0.117)	(0.125)	(0.126)	(0.123)	(0.110)	(0.132)	(0.129)
	Observations	161	173	173	121	164	181	176	143
French 2011	Estimated effect	0.069	-0.078	-0.102	0.082	0.006	-0.147	-0.141	0.119
		(0.170)	(0.167)	(0.167)	(0.170)	(0.152)	(0.183)	(0.173)	(0.185)
	Observations	120	136	135	94	127	143	136	118
Mathematics 2011	Estimated effect	0.087	-0.021	-0.287*	0.134	0.039	-0.093	-0.274*	0.149
		(0.116)	(0.175)	(0.145)	(0.180)	(0.121)	(0.186)	(0.158)	(0.167)
	Observations	120	136	135	94	127	143	136	118
Oral 2011	Estimated effect	0.095	0.208	0.060	0.301	0.055	0.311*	0.031	0.236
		(0.159)	(0.186)	(0.198)	(0.242)	(0.154)	(0.186)	(0.199)	(0.248)
	Observations	100	104	107	75	102	111	108	91
Index 2011	Estimated effect	-0.006	-0.014	-0.259	0.232	-0.024	0.020	-0.311	0.272
		(0.163)	(0.208)	(0.186)	(0.233)	(0.154)	(0.225)	(0.199)	(0.218)
	Observations	98	102	104	73	100	107	105	88

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.24: *Estimated ATE and ATET for Subpopulations, Grade 4*

		ATE				ATET			
		Boys	Girls	Rural zones	Urban zones	Boys	Girls	Rural zones	Urban zones
French 2010	Estimated effect	-0.045	0.053	0.047	-0.185*	-0.168**	0.003	-0.000	-0.140
		(0.089)	(0.114)	(0.109)	(0.095)	(0.085)	(0.101)	(0.106)	(0.086)
	Observations	227	162	230	143	236	176	234	173
Mathematics 2010	Estimated effect	-0.199**	-0.051	-0.116	-0.109	-0.192**	-0.046	-0.182	-0.081
		(0.095)	(0.119)	(0.112)	(0.119)	(0.096)	(0.121)	(0.113)	(0.099)
	Observations	227	162	230	143	236	176	234	173
Oral 2010	Estimated effect	0.045	-0.279**	-0.070	-0.154	-0.041	-0.219*	-0.001	-0.334***
		(0.114)	(0.129)	(0.127)	(0.128)	(0.111)	(0.130)	(0.123)	(0.121)
	Observations	198	144	210	125	206	157	214	150
Index 2010	Estimated effect	-0.078	-0.123	-0.020	-0.180*	-0.147	-0.132	-0.040	-0.196**
		(0.096)	(0.113)	(0.111)	(0.104)	(0.096)	(0.113)	(0.107)	(0.087)
	Observations	197	142	206	124	205	155	210	149
French 2011	Estimated effect	-0.105	-0.160	0.013	-0.490**	-0.237	-0.077	-0.044	-0.319
		(0.132)	(0.179)	(0.124)	(0.222)	(0.149)	(0.178)	(0.129)	(0.209)
	Observations	134	110	134	101	140	122	135	122
Mathematics 2011	Estimated effect	-0.171	-0.125	-0.093	-0.189	-0.328**	-0.023	-0.184	-0.059
		(0.144)	(0.184)	(0.153)	(0.249)	(0.151)	(0.180)	(0.153)	(0.228)
	Observations	134	110	134	101	140	122	135	122
Oral 2011	Estimated effect	-0.213	-0.555***	-0.091	-0.951***	-0.439**	-0.525***	-0.082	-0.931***
		(0.167)	(0.181)	(0.189)	(0.196)	(0.187)	(0.188)	(0.183)	(0.213)
	Observations	112	97	121	82	116	106	123	97
Index 2011	Estimated effect	-0.173	-0.379**	0.014	-0.584***	-0.433**	-0.347**	-0.052	-0.556***
		(0.163)	(0.178)	(0.167)	(0.158)	(0.189)	(0.173)	(0.169)	(0.162)
	Observations	105	91	113	78	109	100	114	93

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.25: *Estimated ATE and ATET by Social Class Level and by Grade*

		Grade 2				Grade 4			
		ATE		ATET		ATE		ATET	
		Poor	Non Poor	Poor	Non Poor	Poor	Non Poor	Poor	Non Poor
French 2010	Estimated effect	-0.182	0.111	-0.100	0.099	0.021	-0.052	-0.054	-0.135*
		(0.160)	(0.093)	(0.160)	(0.101)	(0.113)	(0.079)	(0.114)	(0.074)
	Observations	128	240	130	264	144	251	150	263
Mathematics 2010	Estimated effect	-0.259*	0.036	-0.241*	0.017	-0.154	-0.143*	-0.224*	-0.184**
		(0.138)	(0.080)	(0.144)	(0.080)	(0.117)	(0.085)	(0.124)	(0.081)
	Observations	128	240	130	264	144	251	150	263
Oral 2010	Estimated effect	-0.136	0.086	0.015	0.068	-0.095	-0.109	-0.011	-0.193*
		(0.134)	(0.128)	(0.138)	(0.126)	(0.174)	(0.092)	(0.186)	(0.100)
	Observations	109	207	112	225	127	224	131	235
Index 2010	Estimated effect	-0.248*	0.119	-0.179	0.079	-0.011	-0.119	-0.023	-0.185**
		(0.144)	(0.101)	(0.144)	(0.103)	(0.129)	(0.078)	(0.135)	(0.074)
	Observations	108	203	110	221	126	222	130	233
French 2011	Estimated effect	-0.093	0.073	-0.080	0.192	-0.600**	-0.120	-0.649**	-0.128
		(0.282)	(0.143)	(0.263)	(0.134)	(0.267)	(0.112)	(0.291)	(0.132)
	Observations	83	152	86	173	76	174	80	183
Mathematics 2011	Estimated effect	-0.457	-0.097	-0.390	0.069	-0.553**	-0.167	-0.542**	-0.224
		(0.281)	(0.112)	(0.273)	(0.108)	(0.258)	(0.124)	(0.252)	(0.145)
	Observations	83	152	86	173	76	174	80	183
Oral 2011	Estimated effect	0.141	0.153	0.166	0.149	0.137	-0.190*	0.123	-0.282**
		(0.331)	(0.161)	(0.271)	(0.156)	(0.350)	(0.111)	(0.410)	(0.116)
	Observations	64	123	67	137	71	149	72	156
Index 2011	Estimated effect	-0.180	0.069	-0.190	0.172	-0.222	-0.148	-0.236	-0.264***
		(0.357)	(0.146)	(0.308)	(0.147)	(0.273)	(0.092)	(0.315)	(0.100)
	Observations	62	119	65	133	65	142	66	149

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors accounting for clustering in parentheses.

Table B.26: *Unconditional Quantile Treatment Effects, Grade 2*

	French 2010	Mathematics 2010	Oral 2010	Index 2010	French 2011	Mathematics 2011	Oral 2011	Index 2011
Quantile 5	0.170	-0.335	-0.012	-0.371**	0.260	-0.448	0.096	-0.251
	(0.386)	(0.243)	(0.098)	(0.179)	(0.296)	(0.456)	(0.329)	(0.490)
Quantile 10	0.147	-0.076	-0.006	-0.288	0.072	0.150	0.290	0.146
	(0.297)	(0.270)	(0.148)	(0.261)	(0.316)	(0.508)	(0.373)	(0.459)
Quantile 15	0.053	0.014	-0.089	-0.004	-0.054	-0.034	0.483	0.116
	(0.246)	(0.232)	(0.163)	(0.295)	(0.355)	(0.467)	(0.381)	(0.412)
Quantile 20	0.153	0.169	-0.014	0.017	-0.189	0.203	0.334	0.147
	(0.260)	(0.225)	(0.177)	(0.257)	(0.398)	(0.421)	(0.342)	(0.384)
Quantile 25	0.161	0.169	0.211	0.028	0.094	0.000	0.458	-0.053
	(0.242)	(0.217)	(0.217)	(0.231)	(0.419)	(0.364)	(0.320)	(0.394)
Quantile 30	0.071	0.147	0.185	-0.011	0.189	0.001	0.084	0.267
	(0.199)	(0.213)	(0.226)	(0.212)	(0.409)	(0.327)	(0.332)	(0.417)
Quantile 35	0.000	0.331	0.003	0.017	0.288	0.118	0.189	0.140
	(0.184)	(0.209)	(0.224)	(0.194)	(0.384)	(0.297)	(0.354)	(0.413)
Quantile 40	0.022	0.116	0.152	-0.084	0.094	0.016	0.111	0.058
	(0.202)	(0.188)	(0.221)	(0.204)	(0.345)	(0.256)	(0.368)	(0.386)
Quantile 45	0.028	0.138	0.266	-0.050	0.094	0.135	0.316	0.017
	(0.247)	(0.166)	(0.231)	(0.257)	(0.296)	(0.217)	(0.361)	(0.358)
Quantile 50	0.093	0.138	0.289	0.192	0.000	0.100	0.039	0.199
	(0.248)	(0.159)	(0.238)	(0.299)	(0.242)	(0.194)	(0.340)	(0.325)
Quantile 55	0.203	0.104	0.275	0.249	0.000	0.011	0.115	0.286
	(0.215)	(0.158)	(0.240)	(0.273)	(0.190)	(0.172)	(0.304)	(0.278)
Quantile 60	0.093	0.086	0.256	0.167	0.000	0.009	0.196	0.249
	(0.196)	(0.162)	(0.236)	(0.220)	(0.154)	(0.156)	(0.258)	(0.236)
Quantile 65	0.187	0.145	0.247	0.235	-0.094	0.068	0.210	0.212
	(0.163)	(0.158)	(0.246)	(0.185)	(0.144)	(0.143)	(0.210)	(0.209)
Quantile 70	0.156	0.096	0.059	0.177	0.000	-0.068	0.151	0.117
	(0.128)	(0.147)	(0.253)	(0.175)	(0.137)	(0.123)	(0.186)	(0.189)
Quantile 75	0.111	-0.011	-0.064	0.122	0.073	0.000	0.102	0.083
	(0.129)	(0.133)	(0.249)	(0.178)	(0.123)	(0.100)	(0.184)	(0.177)
Quantile 80	-0.051	-0.110	0.021	0.006	0.000	0.068	0.118	0.033
	(0.130)	(0.120)	(0.222)	(0.165)	(0.103)	(0.093)	(0.180)	(0.165)
Quantile 85	0.039	-0.068	0.059	-0.056	-0.094	0.068	0.020	-0.031
	(0.125)	(0.113)	(0.178)	(0.133)	(0.091)	(0.086)	(0.164)	(0.131)
Quantile 90	-0.009	-0.006	-0.004	-0.077	-0.094	-0.002	0.024	-0.069
	(0.110)	(0.119)	(0.148)	(0.103)	(0.084)	(0.065)	(0.151)	(0.095)
Quantile 95	0.000	-0.068	-0.165	-0.116	0.000	0.000	-0.006	-0.012
	(0.090)	(0.104)	(0.165)	(0.099)	(0.061)	(0.062)	(0.129)	(0.098)
Observations	710	710	615	603	461	461	376	366

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Bootstrap standard errors accounting for clustering in parentheses.

Table B.27: *Unconditional Quantile Treatment Effects, Grade 4*

	French 2010	Mathematics 2010	Oral 2010	Index 2010	French 2011	Mathematics 2011	Oral 2011	Index 2011
Quantile 5	-0.006	-0.035	0.062	-0.032	-0.565	-0.807	-0.906	-0.897
	(0.128)	(0.202)	(0.199)	(0.189)	(0.645)	(0.575)	(0.618)	(0.738)
Quantile 10	0.039	-0.146	-0.085	0.091	-0.124	-0.345	-0.757	-0.238
	(0.143)	(0.195)	(0.258)	(0.210)	(0.373)	(0.567)	(0.728)	(0.558)
Quantile 15	0.058	-0.132	-0.185	-0.027	-0.198	-0.066	-0.577	0.015
	(0.165)	(0.213)	(0.339)	(0.260)	(0.346)	(0.391)	(0.616)	(0.454)
Quantile 20	-0.078	-0.084	-0.466	-0.152	-0.120	0.011	-0.699	-0.091
	(0.224)	(0.215)	(0.336)	(0.256)	(0.337)	(0.239)	(0.457)	(0.385)
Quantile 25	-0.117	-0.048	-0.557**	-0.365	-0.149	0.044	-0.583*	-0.086
	(0.222)	(0.212)	(0.263)	(0.240)	(0.299)	(0.229)	(0.354)	(0.339)
Quantile 30	-0.184	-0.075	-0.521**	-0.427*	-0.154	-0.261	-0.531*	-0.235
	(0.195)	(0.212)	(0.245)	(0.223)	(0.286)	(0.253)	(0.295)	(0.323)
Quantile 35	-0.334*	-0.201	-0.511**	-0.411*	-0.126	-0.318	-0.483*	-0.107
	(0.182)	(0.205)	(0.260)	(0.228)	(0.287)	(0.246)	(0.269)	(0.304)
Quantile 40	-0.234	-0.287	-0.397	-0.250	-0.097	-0.263	-0.425*	-0.276
	(0.173)	(0.196)	(0.272)	(0.216)	(0.269)	(0.230)	(0.250)	(0.272)
Quantile 45	-0.233	-0.250	-0.088	-0.378*	-0.169	-0.277	-0.266	-0.252
	(0.179)	(0.208)	(0.239)	(0.206)	(0.233)	(0.234)	(0.229)	(0.247)
Quantile 50	-0.114	-0.255	-0.036	-0.361*	-0.313	-0.297	-0.319	-0.373
	(0.192)	(0.208)	(0.165)	(0.202)	(0.209)	(0.260)	(0.218)	(0.243)
Quantile 55	-0.033	-0.266	-0.100	-0.261	-0.344*	-0.416	-0.201	-0.285
	(0.200)	(0.180)	(0.116)	(0.178)	(0.209)	(0.265)	(0.215)	(0.238)
Quantile 60	-0.155	-0.190	-0.098	-0.255	-0.219	-0.306	-0.261	-0.292
	(0.200)	(0.143)	(0.104)	(0.163)	(0.211)	(0.234)	(0.211)	(0.227)
Quantile 65	-0.114	-0.248*	-0.139	-0.243	-0.140	-0.302	-0.264	-0.144
	(0.202)	(0.149)	(0.098)	(0.161)	(0.210)	(0.198)	(0.195)	(0.215)
Quantile 70	-0.078	-0.262	-0.144*	-0.167	-0.140	-0.209	-0.252	-0.263
	(0.194)	(0.165)	(0.086)	(0.140)	(0.214)	(0.191)	(0.172)	(0.216)
Quantile 75	0.000	-0.121	-0.150	-0.110	-0.070	-0.223	-0.157	-0.280
	(0.156)	(0.172)	(0.093)	(0.116)	(0.216)	(0.228)	(0.141)	(0.231)
Quantile 80	0.020	-0.270	-0.061	0.009	0.000	-0.199	-0.092	-0.111
	(0.166)	(0.219)	(0.109)	(0.130)	(0.213)	(0.269)	(0.114)	(0.247)
Quantile 85	0.000	-0.314	-0.043	-0.059	0.071	-0.047	-0.130	0.012
	(0.198)	(0.234)	(0.099)	(0.178)	(0.191)	(0.260)	(0.109)	(0.244)
Quantile 90	0.055	-0.013	-0.032	-0.119	0.070	0.115	-0.046	0.163
	(0.204)	(0.232)	(0.100)	(0.210)	(0.156)	(0.207)	(0.117)	(0.203)
Quantile 95	-0.155	-0.289	-0.037	-0.106	0.000	0.136	-0.073	-0.079
	(0.184)	(0.234)	(0.106)	(0.210)	(0.124)	(0.142)	(0.105)	(0.121)
Observations	717	717	638	633	453	453	398	378

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Bootstrap standard errors accounting for clustering in parentheses.

Table B.28: *Unconditional Quantile Treatment Effects on the Treated, Grade 2*

	French 2010	Mathematics 2010	Oral 2010	Index 2010	French 2011	Mathematics 2011	Oral 2011	Index 2011
Quantile 5	0.040	-0.344	-0.012	-0.253	0.156	-0.586	-0.043	-0.297
	(0.304)	(0.212)	(0.131)	(0.180)	(0.343)	(0.496)	(0.322)	(0.511)
Quantile 10	0.000	-0.295	-0.098	-0.275	0.046	-0.152	0.236	-0.079
	(0.204)	(0.204)	(0.170)	(0.275)	(0.349)	(0.368)	(0.323)	(0.417)
Quantile 15	-0.156	-0.161	-0.092	-0.043	-0.323	-0.169	0.368	-0.238
	(0.229)	(0.222)	(0.153)	(0.237)	(0.358)	(0.342)	(0.360)	(0.404)
Quantile 20	-0.089	0.037	-0.099	-0.060	-0.189	-0.161	0.228	-0.246
	(0.234)	(0.212)	(0.212)	(0.257)	(0.426)	(0.348)	(0.376)	(0.447)
Quantile 25	0.000	-0.069	0.076	-0.159	-0.086	-0.068	-0.033	-0.290
	(0.209)	(0.193)	(0.264)	(0.266)	(0.480)	(0.311)	(0.385)	(0.476)
Quantile 30	-0.140	-0.066	-0.109	-0.134	0.094	-0.119	-0.173	-0.213
	(0.216)	(0.216)	(0.256)	(0.231)	(0.432)	(0.271)	(0.415)	(0.453)
Quantile 35	-0.087	0.127	-0.123	-0.194	0.000	-0.135	-0.006	-0.017
	(0.243)	(0.230)	(0.252)	(0.233)	(0.367)	(0.231)	(0.429)	(0.409)
Quantile 40	-0.247	0.026	-0.076	-0.246	0.094	0.032	-0.198	-0.065
	(0.247)	(0.221)	(0.277)	(0.273)	(0.294)	(0.200)	(0.440)	(0.362)
Quantile 45	-0.010	0.058	-0.051	-0.336	0.000	0.068	-0.130	-0.110
	(0.238)	(0.220)	(0.308)	(0.312)	(0.249)	(0.193)	(0.437)	(0.332)
Quantile 50	0.091	0.026	-0.094	-0.138	0.000	0.000	-0.200	0.068
	(0.216)	(0.213)	(0.332)	(0.306)	(0.201)	(0.181)	(0.383)	(0.298)
Quantile 55	-0.053	-0.052	0.055	0.060	0.013	-0.126	-0.127	0.191
	(0.207)	(0.215)	(0.362)	(0.246)	(0.170)	(0.170)	(0.319)	(0.262)
Quantile 60	0.073	-0.036	-0.188	0.001	-0.048	0.000	-0.006	0.167
	(0.188)	(0.220)	(0.347)	(0.225)	(0.151)	(0.167)	(0.260)	(0.238)
Quantile 65	0.169	0.000	-0.257	-0.011	0.000	-0.135	0.053	-0.044
	(0.165)	(0.209)	(0.303)	(0.237)	(0.150)	(0.153)	(0.223)	(0.234)
Quantile 70	0.018	-0.174	-0.270	-0.047	0.000	-0.126	-0.046	0.033
	(0.179)	(0.168)	(0.250)	(0.240)	(0.148)	(0.129)	(0.211)	(0.232)
Quantile 75	0.000	-0.091	-0.108	-0.176	-0.094	-0.068	-0.076	-0.212
	(0.183)	(0.127)	(0.224)	(0.209)	(0.128)	(0.106)	(0.205)	(0.212)
Quantile 80	-0.121	-0.078	-0.013	-0.175	-0.094	-0.064	-0.120	-0.164
	(0.163)	(0.108)	(0.175)	(0.159)	(0.111)	(0.091)	(0.180)	(0.178)
Quantile 85	-0.104	-0.068	0.051	-0.147	-0.094	-0.068	-0.166	-0.113
	(0.129)	(0.123)	(0.160)	(0.108)	(0.094)	(0.074)	(0.148)	(0.119)
Quantile 90	-0.093	-0.054	-0.099	-0.096	0.000	0.000	-0.046	-0.114
	(0.102)	(0.129)	(0.148)	(0.088)	(0.085)	(0.062)	(0.122)	(0.073)
Quantile 95	-0.093	-0.075	-0.154	-0.110	-0.094*	0.000	0.000	-0.051
	(0.068)	(0.115)	(0.162)	(0.098)	(0.056)	(0.054)	(0.095)	(0.080)
Observations	710	710	615	603	461	461	376	366

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Bootstrap standard errors accounting for clustering in parentheses.

Table B.29: *Unconditional Quantile Treatment Effects on the Treated, Grade 4*

	French 2010	Mathematics 2010	Oral 2010	Index 2010	French 2011	Mathematics 2011	Oral 2011	Index 2011
Quantile 5	-0.027	0.062	0.092	0.030	-1.399*	-1.022*	-0.962	-0.998
	(0.114)	(0.219)	(0.214)	(0.192)	(0.769)	(0.571)	(0.650)	(0.630)
Quantile 10	0.058	0.026	-0.016	0.260	-0.304	-0.631	-0.821	-0.605
	(0.174)	(0.224)	(0.405)	(0.257)	(0.420)	(0.486)	(0.729)	(0.570)
Quantile 15	0.075	-0.104	-0.258	-0.181	-0.481	-0.264	-1.000*	-0.334
	(0.294)	(0.207)	(0.482)	(0.380)	(0.307)	(0.341)	(0.529)	(0.447)
Quantile 20	-0.145	0.034	-0.533	-0.346	-0.282	-0.120	-0.915**	-0.341
	(0.335)	(0.226)	(0.379)	(0.357)	(0.318)	(0.288)	(0.418)	(0.407)
Quantile 25	-0.259	0.027	-0.428	-0.412	-0.280	-0.326	-0.564	-0.681*
	(0.274)	(0.268)	(0.313)	(0.308)	(0.309)	(0.282)	(0.345)	(0.396)
Quantile 30	-0.407*	-0.203	-0.491*	-0.487	-0.443	-0.522*	-0.653**	-0.627*
	(0.223)	(0.270)	(0.275)	(0.303)	(0.315)	(0.271)	(0.309)	(0.347)
Quantile 35	-0.288	-0.163	-0.333	-0.330	-0.478	-0.439	-0.718**	-0.736***
	(0.195)	(0.262)	(0.284)	(0.289)	(0.308)	(0.278)	(0.290)	(0.276)
Quantile 40	-0.196	-0.293	-0.175	-0.442*	-0.540**	-0.412	-0.611**	-0.686***
	(0.199)	(0.258)	(0.256)	(0.257)	(0.261)	(0.284)	(0.255)	(0.215)
Quantile 45	-0.057	-0.289	-0.176	-0.266	-0.490**	-0.655**	-0.628***	-0.689***
	(0.238)	(0.238)	(0.177)	(0.224)	(0.209)	(0.270)	(0.214)	(0.197)
Quantile 50	-0.078	-0.216	-0.202	-0.270	-0.490***	-0.558**	-0.671***	-0.628***
	(0.275)	(0.192)	(0.127)	(0.206)	(0.181)	(0.245)	(0.179)	(0.215)
Quantile 55	-0.188	-0.216	-0.138	-0.261	-0.371**	-0.594***	-0.528***	-0.484**
	(0.267)	(0.149)	(0.123)	(0.205)	(0.185)	(0.219)	(0.147)	(0.237)
Quantile 60	-0.261	-0.149	-0.172	-0.228	-0.335*	-0.457**	-0.481***	-0.521**
	(0.233)	(0.162)	(0.115)	(0.193)	(0.196)	(0.221)	(0.123)	(0.243)
Quantile 65	-0.209	-0.216	-0.173*	-0.210	-0.420*	-0.390*	-0.410***	-0.486**
	(0.195)	(0.189)	(0.095)	(0.154)	(0.218)	(0.235)	(0.110)	(0.235)
Quantile 70	-0.078	-0.169	-0.202**	-0.115	-0.469*	-0.409	-0.403***	-0.495**
	(0.179)	(0.225)	(0.096)	(0.138)	(0.244)	(0.253)	(0.113)	(0.225)
Quantile 75	-0.045	-0.122	-0.092	-0.061	-0.420*	-0.429	-0.297***	-0.491**
	(0.201)	(0.274)	(0.109)	(0.173)	(0.254)	(0.277)	(0.111)	(0.217)
Quantile 80	-0.078	-0.377	-0.150	-0.052	-0.303	-0.265	-0.300***	-0.489**
	(0.255)	(0.296)	(0.105)	(0.230)	(0.213)	(0.234)	(0.108)	(0.228)
Quantile 85	-0.155	-0.325	-0.121	-0.130	-0.280	-0.205	-0.225**	-0.271
	(0.289)	(0.280)	(0.088)	(0.271)	(0.180)	(0.174)	(0.107)	(0.228)
Quantile 90	-0.233	0.000	-0.021	-0.213	-0.210	0.055	-0.278**	-0.168
	(0.261)	(0.281)	(0.102)	(0.271)	(0.151)	(0.178)	(0.109)	(0.163)
Quantile 95	-0.085	-0.284	-0.053	-0.256	-0.070	-0.065	-0.162**	-0.176*
	(0.184)	(0.216)	(0.090)	(0.216)	(0.116)	(0.157)	(0.077)	(0.093)
Observations	717	717	638	633	453	453	398	378

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Bootstrap standard errors accounting for clustering in parentheses.

Figure B.2: *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test*

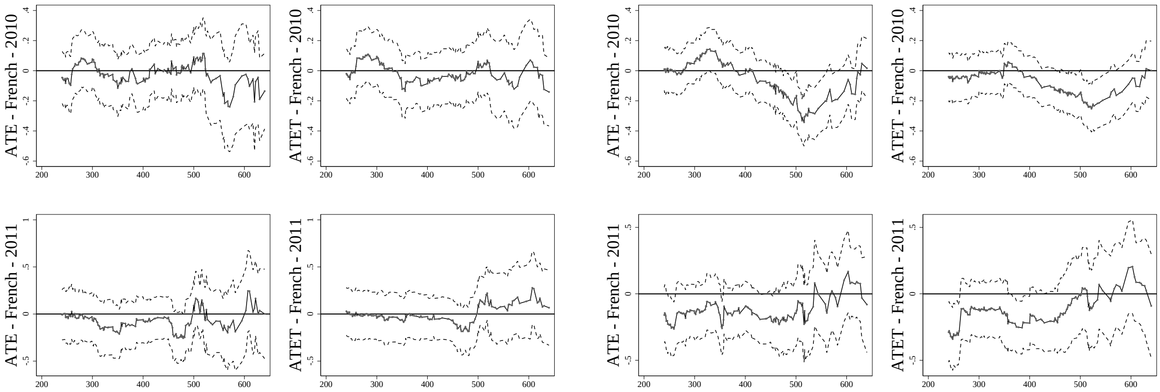


Figure B.3: *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test*

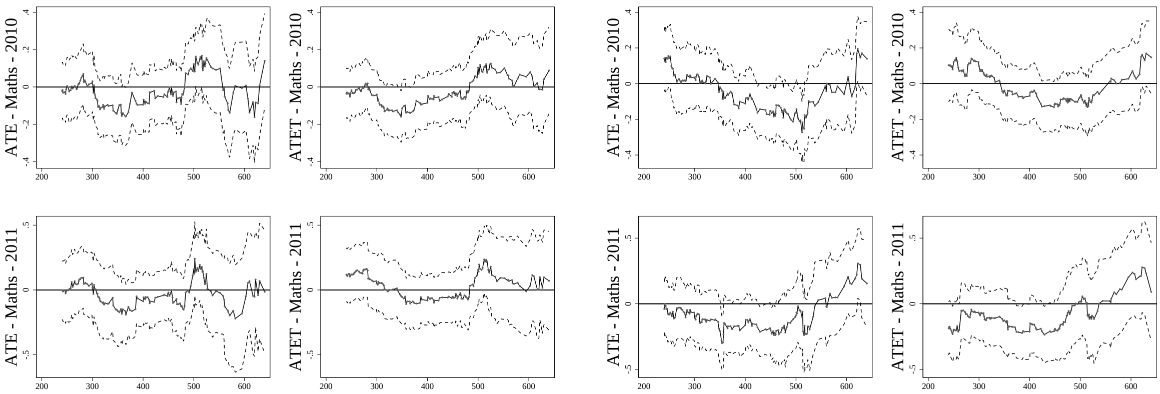


Figure B.4: *Sensitivity of the Estimates to the Selection of the Cut-off Point, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test*

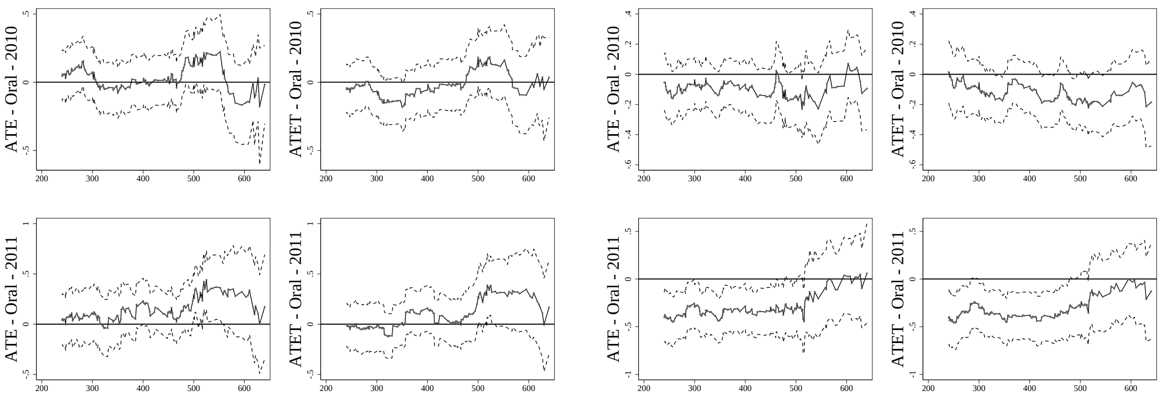


Figure B.5: *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables to the Set of Regressors, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test*

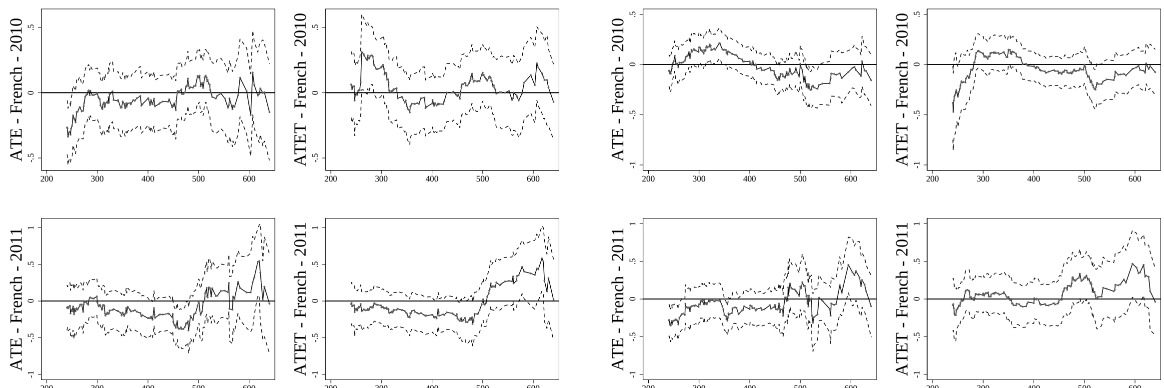


Figure B.6: *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables to the Set of Regressors, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test*

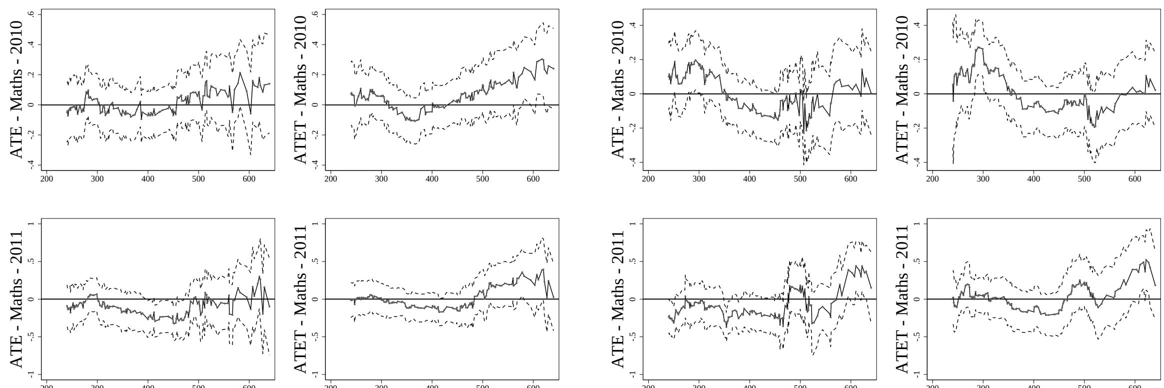


Figure B.7: *Sensitivity of the Estimates to the Inclusion of Size-dependent Variables to the Set of Regressors, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test*

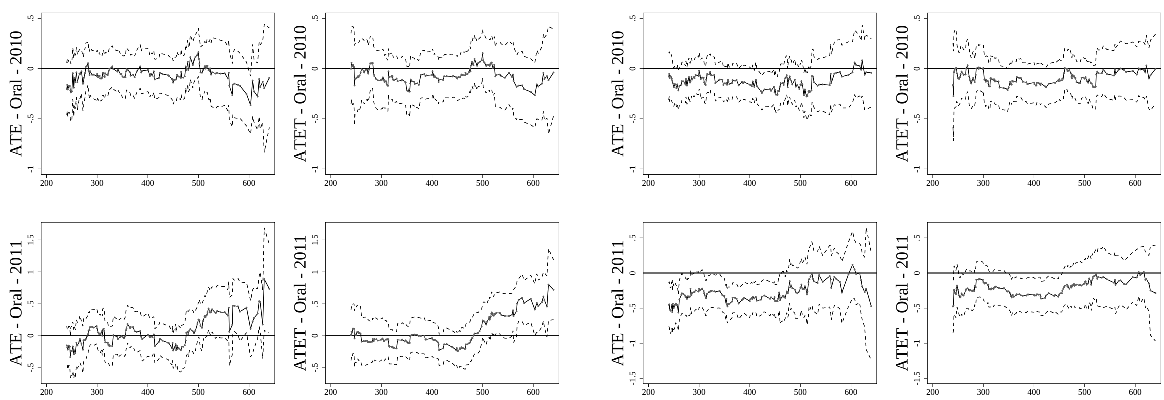


Figure B.8: *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test*

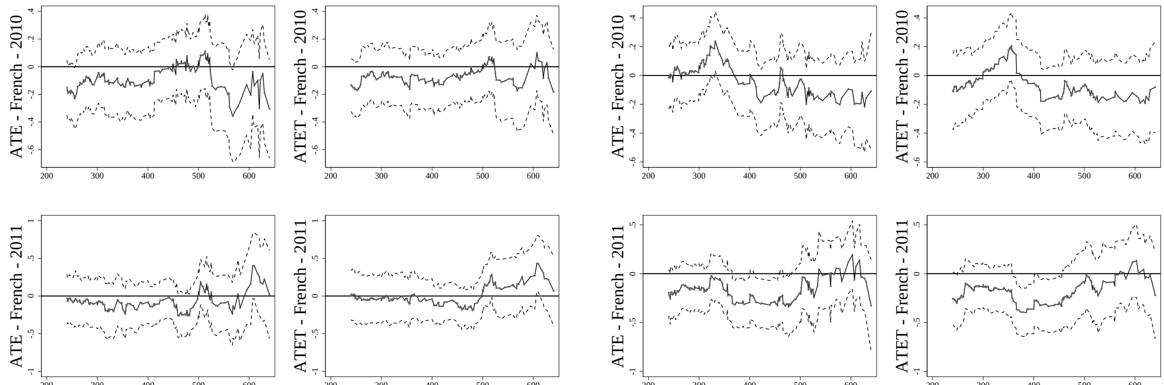


Figure B.9: *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test*

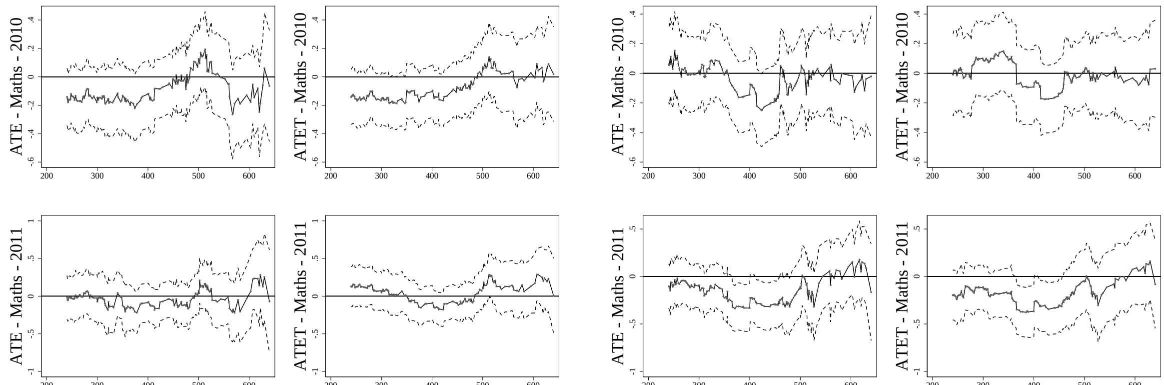


Figure B.10: *Sensitivity of the Estimates to the Exclusion of Prior Ability of Students, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test*

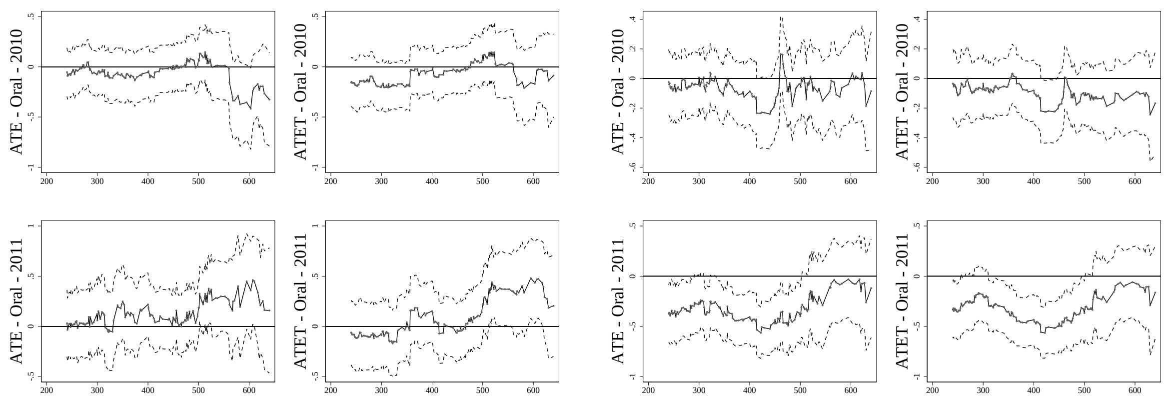


Figure B.11: *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test*

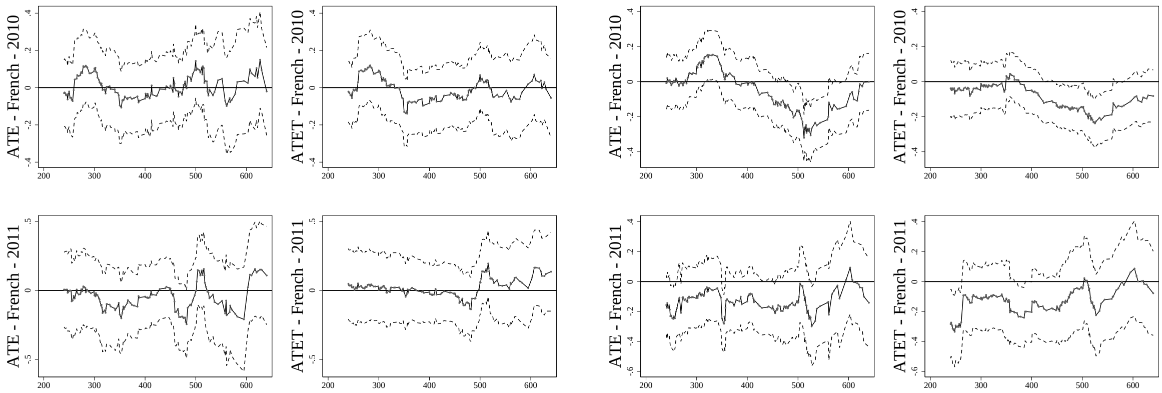


Figure B.12: *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test*

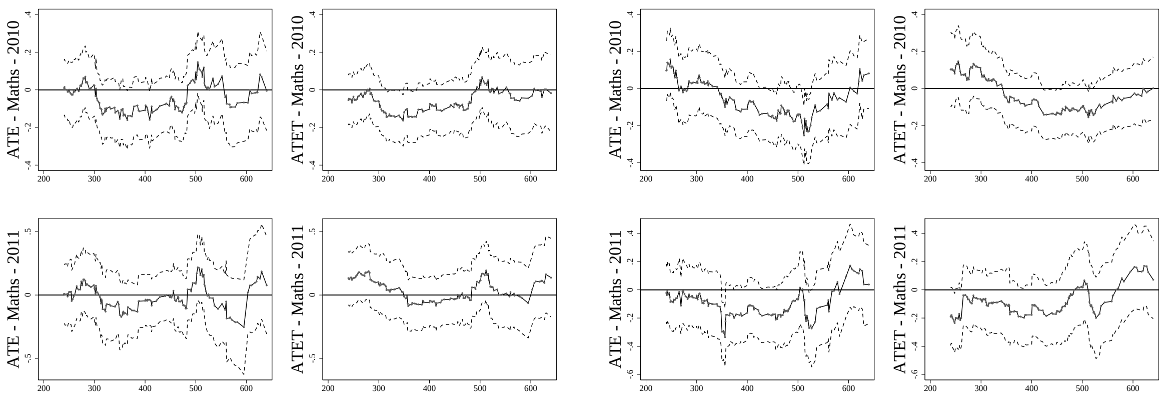


Figure B.13: *Sensitivity of the Estimates to the Inclusion of Too Large Schools, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test*

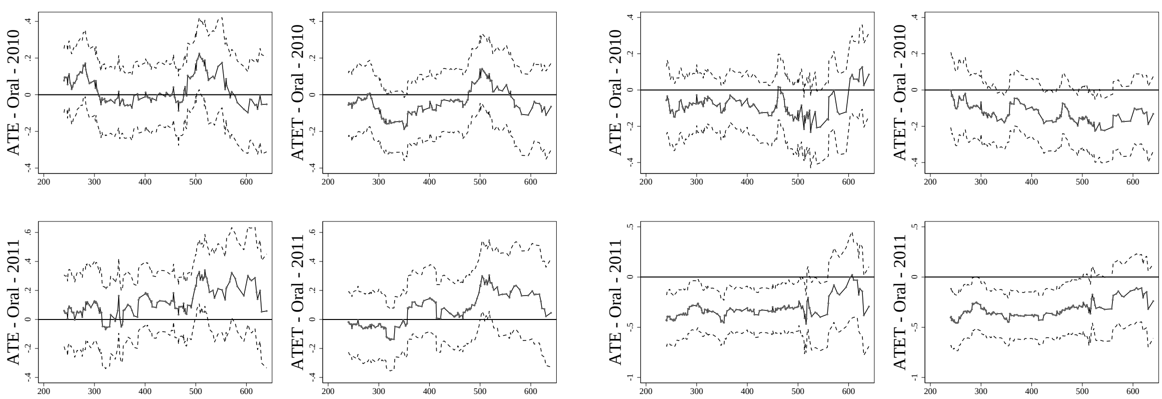


Figure B.14: *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test*

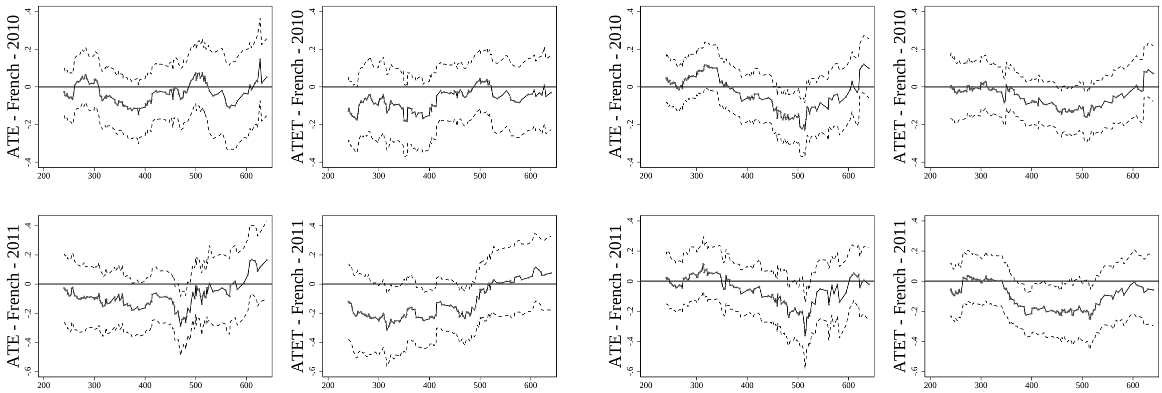


Figure B.15: *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test*

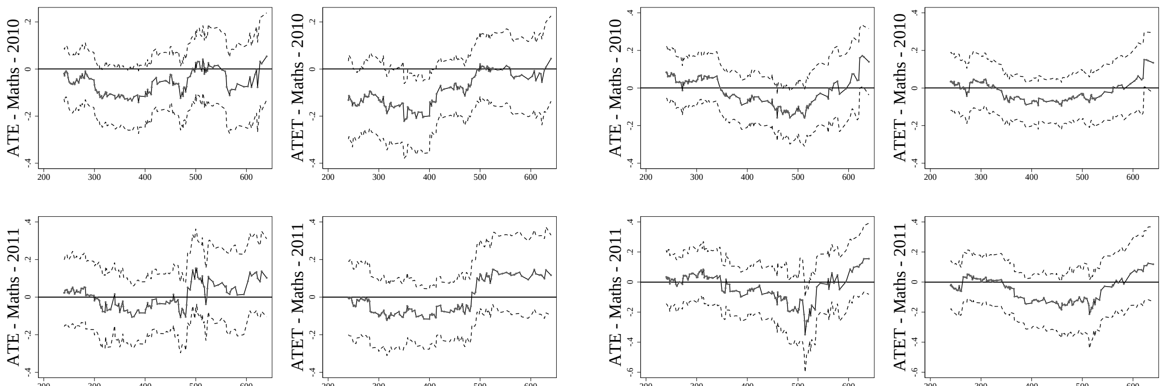


Figure B.16: *Sensitivity of the Estimates to the Inclusion of Students with no Household Data, ATE and ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test*

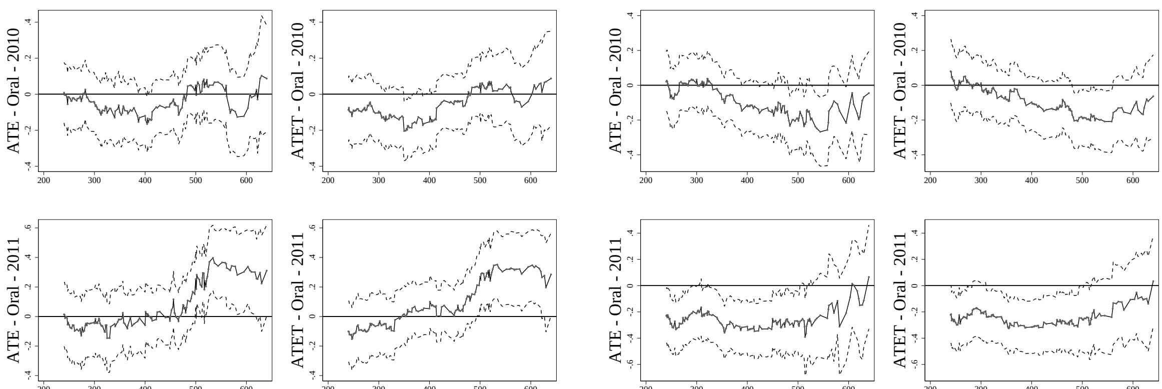


Figure B.17: *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), French Test*

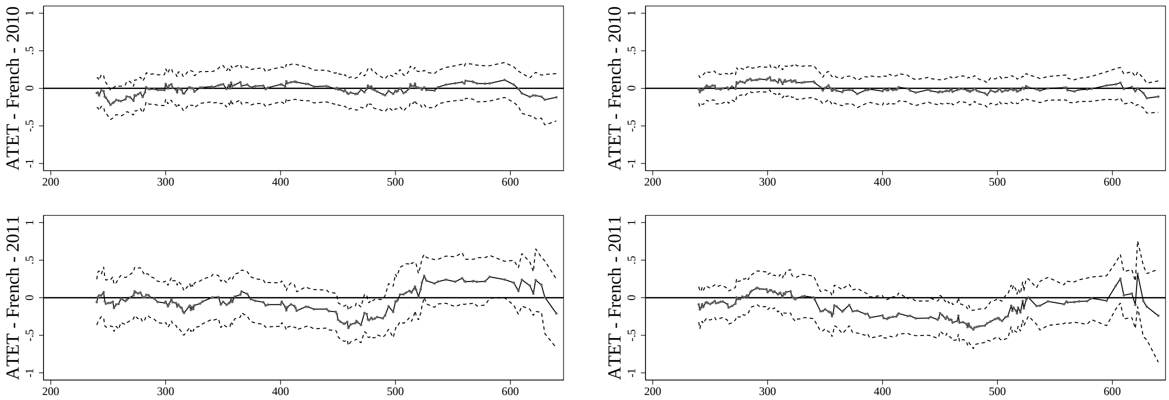


Figure B.18: *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Mathematics Test*

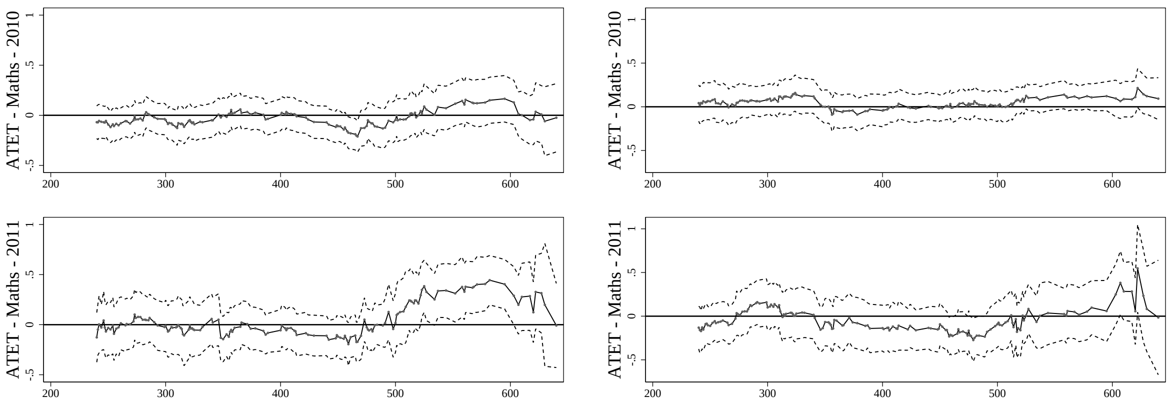
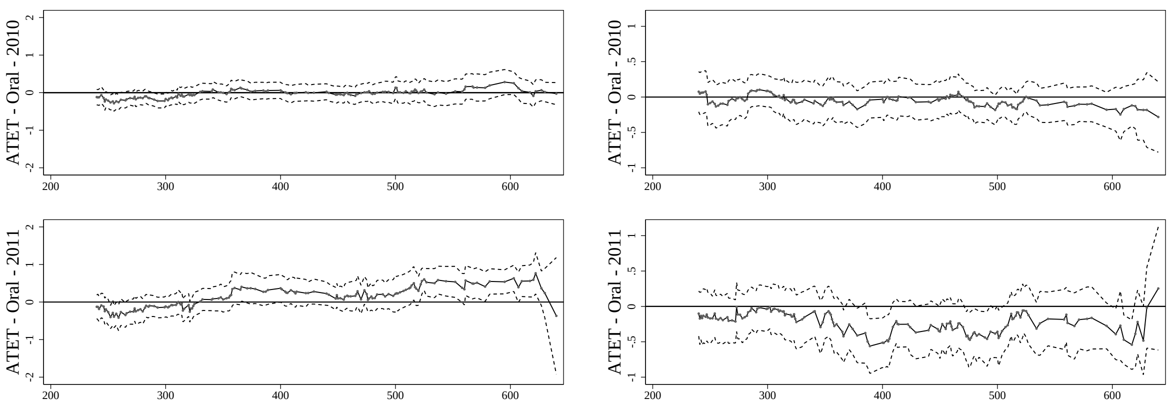


Figure B.19: *Sensitivity of the Estimates to the Restriction of the Sample to the Thick Support, ATET within Confidence Interval, Grade 2 (left side) and Grade 4 (right side), Oral Test*



Appendix C

Supplement to Chapter 3

Table C.1: *Descriptive Statistics by Country*

	Cameroon	Chad	Congo	Ivory Coast	Senegal	Togo
Panel A : Test scores						
Average score in French (out of 100) - Grade 2	66.5 (22.2)	39.1 (23.6)	44.3 (25.9)	38.5 (24.5)	45.0 (23.3)	34.1 (23.1)
Average score in French (out of 100) - Grade 5	45.8 (18.0)	38.0 (22.6)	34.2 (20.0)	33.2 (17.1)	38.3 (15.6)	29.1 (15.0)
Average score in mathematics (out of 100) - Grade 2	55.8 (23.8)	40.3 (26.2)	45.2 (26.0)	27.6 (20.6)	47.2 (22.6)	38.6 (25.1)
Average score in mathematics (out of 100) - Grade 5	46.4 (17.1)	38.1 (20.6)	35.8 (17.4)	27.8 (12.4)	41.8 (14.8)	33.7 (14.9)
Panel B: Schools and principals characteristics						
Age of the principal (in years)	42.330 (7.938)	41.390 (10.040)	45.510 (7.010)	44.440 (6.655)	49.490 (8.159)	47.660 (6.193)
Experience of the principal (in years)	10.260 (8.042)	6.708 (5.133)	8.529 (6.003)	7.347 (5.666)	11.730 (8.495)	10.250 (7.751)
Percentage of students with a female principal	0.135 (0.342)	0.094 (0.292)				
Percentage of students attending a school led by a principal with the BEPC[1] or an equivalent diploma	0.474 (0.499)				0.615 (0.487)	0.635 (0.482)
Average school size	351.200 (264.500)	176.500 (157.600)	707.000 (565.800)	307.400 (111.500)	603.200 (369.900)	118.400 (52.350)
Percentage of students attending a school located in a rural region	0.512 (0.500)	0.599 (0.490)	0.285 (0.452)	0.536 (0.499)	0.428 (0.495)	0.631 (0.482)
Percentage of students attending a school equipped with a functional library in the school	0.110 (0.313)	0.0189 (0.136)	0.0544 (0.227)	0.0915 (0.288)	0.277 (0.448)	0.0300 (0.170)
Percentage of students attending an electrified school	0.200 (0.400)	0.084 (0.277)	0.187 (0.390)	0.411 (0.492)	0.566 (0.496)	0.360 (0.480)
Percentage of students attending a school with piped water in the school	0.341 (0.474)	0.327 (0.469)	0.396 (0.489)	0.529 (0.499)	0.842 (0.365)	0.185 (0.388)
Percentage of students attending a school with a management committee	0.127 (0.334)		0.599 (0.490)		0.713 (0.453)	
Panel C: Classrooms and teachers characteristics						
Percentage of students with a female teacher	0.321 (0.467)	0.195 (0.396)	0.516 (0.500)	0.341 (0.474)	0.362 (0.481)	0.506 (0.500)
Average experience of the teachers (in years)	10.340 (8.116)		9.910 (7.903)	9.380 (8.156)	9.921 (7.622)	10.820 (8.201)
Average age of the teachers (in years)	35.960 (8.059)	33.200 (7.584)	37.800 (8.326)	35.230 (7.828)	35.850 (7.668)	37.710 (9.455)
Percentage of students with a teacher who holds the BEPC or an equivalent diploma	0.503 (0.500)			0.537 (0.499)		0.586 (0.493)

Table C.1: *Continued*

	Cameroon	Chad	Congo	Ivory Coast	Senegal	Togo
Percentage of students in classroom with less than 40 students	0.227	0.347	0.265	0.305	0.241	0.396
	(0.419)	(0.476)	(0.442)	(0.460)	(0.428)	(0.489)
Percentage of students with a teacher who uses the reading manual in teaching	0.722	0.703	0.838	0.863	0.875	0.908
	(0.448)	(0.457)	(0.368)	(0.344)	(0.330)	(0.289)
Percentage of students with a teacher who uses the mathematics manual in teaching	0.712	0.668	0.819	0.859	0.834	0.842
	(0.453)	(0.471)	(0.385)	(0.348)	(0.372)	(0.365)
Percentage of students attending a multi-grade class				0.066	0.064	
				(0.248)	(0.245)	
Panel D : Students and households characteristics						
Percentage of female students	0.439	0.416		0.471	0.498	0.455
	(0.496)	(0.493)		(0.499)	(0.500)	(0.498)
Average age of grade 2 students (in years)	7.894	8.824	7.523	7.810	8.248	7.710
	(1.508)	(1.503)	(1.197)	(1.342)	(1.116)	(1.400)
Average age of grade 5 students (in years)	11.639	12.613	11.401	11.343	11.677	11.174
	(1.752)	(1.717)	(1.554)	(1.583)	(1.378)	(1.826)
Percentage of students that participate in housework	0.821	0.666	0.673	0.744	0.600	0.852
	(0.383)	(0.472)	(0.469)	(0.436)	(0.490)	(0.355)
Percentage of students that participate in field work	0.643	0.524	0.304	0.571	0.384	0.611
	(0.479)	(0.500)	(0.460)	(0.495)	(0.486)	(0.488)
Percentage of students that participate in commercial activities	0.309	0.226	0.194	0.228	0.150	0.201
	(0.462)	(0.418)	(0.395)	(0.420)	(0.357)	(0.401)
Average number of repetitions	0.820	0.672	0.664	0.562	0.505	0.729
	(0.746)	(0.601)	(0.846)	(0.773)	(0.661)	(0.717)
Percentage of students speaking the French language at home	0.262	0.0745	0.255	0.223	0.0361	0.0725
	(0.440)	(0.263)	(0.436)	(0.416)	(0.187)	(0.259)
Percentage of students with a literate mother	0.618	0.276	0.777	0.452	0.383	
	(0.486)	(0.447)	(0.416)	(0.498)	(0.486)	
Percentage of students that get help with schoolwork at home	0.281		0.372	0.207	0.185	
	(0.450)		(0.483)	(0.405)	(0.388)	
Percentage of students that use the reading book in class	0.541	0.344		0.764	0.623	0.474
	(0.498)	(0.475)		(0.425)	(0.485)	(0.499)
Percentage of students that use the mathematics book in class	0.406	0.289		0.726	0.528	0.377
	(0.491)	(0.453)		(0.446)	(0.499)	(0.485)

[1] BEPC is Brevet d'Etudes du Premier Cycle. It is a secondary school certificate obtained after 10 years of schooling.

Figure C.1: *Correlations of Grades 2 and 5 Test Scores in French and Mathematics*

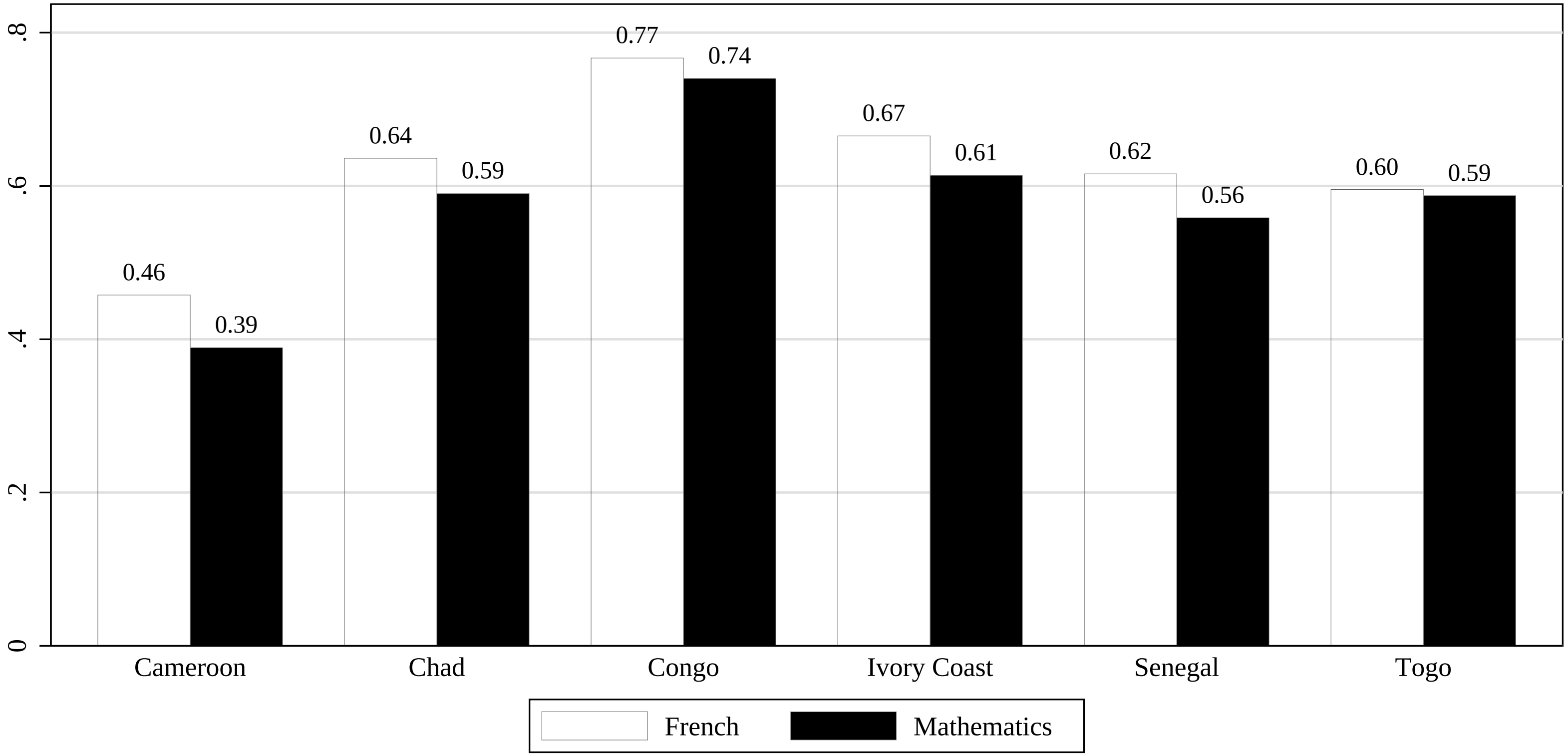


Table C.2: *Empirical χ^2 values derived from likelihood ratio tests comparing 3-level to 2-level models*

		Cameroon	Chad	Congo	Ivory Coast	Senegal	Togo
Analysis of variance (no explanatory variable)	French	290.07	645.12	268.30	177.02	267.65	272.31
	Mathematics	310.52	443.22	212.29	130.80	224.79	189.08

Table C.3: *Reduction in Variance by Type of Conditional Analysis of Variance*

		Socioeconomic status			Initial achievement		
		Reduction at the school-level	Reduction at the classroom-level	Reduction at the student-level	Reduction at the school-level	Reduction at the classroom-level	Reduction at the student-level
Cameroon	French	9.71%	-0.89%	0.25%	52.79%	6.79%	9.79%
	Mathematics	5.27%	0.46%	-0.01%	37.57%	3.08%	9.22%
Chad	French	1.88%	-0.07%	0.12%	57.80%	20.06%	20.93%
	Mathematics	2.24%	0.74%	0.05%	66.80%	3.04%	22.07%
Congo	French	6.35%	0.41%	0.49%	54.08%	-0.66%	17.15%
	Mathematics	5.56%	-0.40%	0.38%	57.19%	-4.49%	15.97%
Ivory Coast	French	15.08%	-1.06%	0.04%	80.92%	38.40%	46.59%
	Mathematics	14.71%	-0.81%	0.13%	83.25%	25.40%	39.64%
Senegal	French	16.01%	-0.28%	0.32%	79.09%	30.25%	40.60%
	Mathematics	14.27%	0.19%	-0.01%	79.31%	25.61%	36.32%
Togo	French	31.61%	-0.28%	0.49%	75.35%	14.19%	39.76%
	Mathematics	18.43%	-0.17%	0.24%	76.35%	-17.51%	36.62%

Table C.4: *Estimated Models for Cameroon*

	French					Mathematics				
	Model 0[2]	Model 1[3]	Model 2[4]	Model 3[5]	Model 4[6]	Model 0	Model 1	Model 2	Model 3	Model 4
Grade dummy (1=Grade 5)		0.119	0.114	0.110	0.056		-0.007	-0.002	-0.020	-0.075
		(0.076)	(0.083)	(0.089)	(0.083)		(0.079)	(0.093)	(0.106)	(0.102)
Female student		-0.036	-0.035	-0.035	-0.005		-0.081**	-0.048	-0.048	-0.010
		(0.030)	(0.039)	(0.039)	(0.037)		(0.030)	(0.038)	(0.038)	(0.036)
Age of the student (in years)		-0.021+	-0.020+	-0.020+	-0.017+		0.009	0.010	0.010	0.012
		(0.011)	(0.011)	(0.011)	(0.010)		(0.011)	(0.011)	(0.011)	(0.011)
Socioeconomic status of the student's family		0.058**	0.045*	0.045*	0.019		0.043+	0.021	0.021	-0.002
		(0.021)	(0.023)	(0.023)	(0.020)		(0.023)	(0.025)	(0.025)	(0.023)
The student gets help with schoolwork at home		-0.018	-0.019	-0.018	-0.003		-0.023	-0.024	-0.023	-0.010
		(0.032)	(0.032)	(0.032)	(0.031)		(0.034)	(0.034)	(0.034)	(0.033)
Number of times the student repeats a grade		-0.086***	-0.086***	-0.084***	-0.067**		-0.049*	-0.049*	-0.049*	-0.035+
		(0.022)	(0.022)	(0.022)	(0.021)		(0.022)	(0.022)	(0.022)	(0.020)
The student's mother is literate		0.046	0.043	0.045	0.029		0.011	0.003	0.006	-0.011
		(0.036)	(0.037)	(0.037)	(0.034)		(0.038)	(0.038)	(0.038)	(0.035)
The student participates in commercial activities		0.001	0.002	0.002	0.010		-0.014	-0.011	-0.011	-0.003
		(0.033)	(0.032)	(0.033)	(0.031)		(0.035)	(0.034)	(0.034)	(0.032)
The student participates in field work		-0.048	-0.044	-0.042	-0.039		0.019	0.025	0.026	0.028
		(0.040)	(0.041)	(0.040)	(0.039)		(0.034)	(0.034)	(0.034)	(0.033)
The student uses the French language at home		0.018	0.011	0.013	-0.028		-0.011	-0.024	-0.022	-0.052
		(0.045)	(0.045)	(0.045)	(0.039)		(0.045)	(0.044)	(0.044)	(0.041)
Female teacher			-0.083	-0.081	-0.071			-0.108	-0.099	-0.099
			(0.094)	(0.096)	(0.088)			(0.091)	(0.092)	(0.085)
Female student x Female teacher			-0.004	-0.003	0.019			-0.099	-0.100	-0.071
			(0.068)	(0.068)	(0.061)			(0.063)	(0.063)	(0.056)
Equipment level of the classroom			0.032	0.023	0.035			0.016	0.014	0.026
			(0.041)	(0.040)	(0.036)			(0.040)	(0.040)	(0.037)
Classroom-level average of the socioeconomic			0.169*	0.093	0.001			0.275***	0.279**	0.185+
status of students' families			(0.080)	(0.095)	(0.073)			(0.082)	(0.102)	(0.095)
Classroom-level standard deviation of			-0.104	-0.040	-0.007			-0.132	-0.105	-0.062
the socioeconomic status			(0.174)	(0.192)	(0.144)			(0.149)	(0.175)	(0.148)
The number of students within the classroom			0.127	0.113	0.118			0.153+	0.147+	0.153+
is less than 40			(0.091)	(0.097)	(0.088)			(0.083)	(0.086)	(0.079)

Table C.4: *Continued*

	French					Mathematics				
	Model 0[2]	Model 1[3]	Model 2[4]	Model 3[4]	Model 4[5]	Model 0	Model 1	Model 2	Model 3	Model 4
The teacher holds a certificate of assistant teachers (CAPIA)			-0.090 (0.116)	-0.038 (0.118)	-0.048 (0.112)			-0.126 (0.122)	-0.106 (0.123)	-0.129 (0.130)
The teacher holds a certificate of teachers of primary and nursery education (CAPIEMP or CAPI)			-0.094 (0.092)	-0.061 (0.096)	-0.066 (0.092)			-0.219* (0.101)	-0.218* (0.105)	-0.217* (0.102)
Experience of the principal (in years)				-0.009 (0.006)	-0.006 (0.005)				-0.008 (0.006)	-0.006 (0.006)
School size (in hundreds of students)				-0.002 (0.020)	0.017 (0.019)				0.008 (0.019)	0.029 (0.019)
Equipment index of the school				0.066 (0.056)	0.041 (0.042)				-0.023 (0.057)	-0.047 (0.050)
The school is located in a rural area				-0.046 (0.110)	-0.127 (0.089)				-0.012 (0.110)	-0.094 (0.103)
Standardized value of initial performance in French test score					0.278*** (0.027)					0.187*** (0.025)
Standardized value of initial performance in mathematics test score					0.125*** (0.026)					0.204*** (0.025)
Intercept	-0.003 (0.052)	0.232* (0.110)	0.346* (0.154)	0.407* (0.182)	0.313* (0.154)	-0.001 (0.050)	-0.016 (0.105)	0.179 (0.143)	0.223 (0.181)	0.139 (0.162)
lns1_1_1_cons[7]	-0.802*** (0.107)	-0.899*** (0.124)	-0.999*** (0.155)	-1.005*** (0.152)	-1.351*** (0.193)	-0.916*** (0.142)	-0.947*** (0.149)	-1.101*** (0.173)	-1.116*** (0.175)	-1.345*** (0.231)
lns2_1_1_cons[8]	-0.827*** (0.082)	-0.817*** (0.083)	-0.797*** (0.084)	-0.808*** (0.084)	-0.862*** (0.078)	-0.781*** (0.114)	-0.783*** (0.112)	-0.794*** (0.108)	-0.794*** (0.109)	-0.817*** (0.114)
lnsig_e _cons[9]	-0.251*** (0.022)	-0.256*** (0.023)	-0.256*** (0.023)	-0.256*** (0.023)	-0.306*** (0.025)	-0.233*** (0.023)	-0.235*** (0.023)	-0.235*** (0.023)	-0.236*** (0.023)	-0.282*** (0.025)
Akaike information criterion	8,372.825	8,346.682	8,354.491	8,358.794	7,998.494	8,482.464	8,485.063	8,475.363	8,481.437	8,161.489
Observations	3,358	3,358	3,358	3,358	3,358	3,358	3,358	3,358	3,358	3,358

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[2]Model 0 is the model for the variance decomposition.

[3]Model 1 is the student-level model.

[4]Model 2 is obtained by adding teachers and classrooms characteristics to Model 1.

[5]This model is obtained by adding principal and school data to Model 2.

[6]This is the final model obtained by inserting students' initial achievement to Model 3.

[7]This is the natural log of the standard deviation of the level-3 errors.

[8]This is the natural log of the standard deviation of the level-2 errors.

[9]This is the natural log of the standard deviation of the level-1 errors.

Table C.5: *Estimated Models for Chad*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Grade dummy (1=Grade 5)		-0.107	-0.072	-0.081	-0.006		-0.180*	-0.154*	-0.154+	-0.066
		(0.076)	(0.084)	(0.084)	(0.072)		(0.072)	(0.077)	(0.079)	(0.071)
Female student		-0.121***	-0.112***	-0.112***	-0.039		-0.204***	-0.192***	-0.192***	-0.106***
		(0.026)	(0.030)	(0.030)	(0.027)		(0.026)	(0.030)	(0.030)	(0.027)
Age of the student (in years)		0.037***	0.037***	0.037***	0.013+		0.051***	0.051***	0.051***	0.024***
		(0.008)	(0.008)	(0.008)	(0.007)		(0.008)	(0.008)	(0.008)	(0.007)
Socioeconomic status of the student's family		0.032+	0.027	0.027	-0.005		0.035+	0.021	0.021	-0.016
		(0.020)	(0.021)	(0.021)	(0.020)		(0.020)	(0.021)	(0.021)	(0.016)
Number of times the student repeats a grade		-0.121***	-0.120***	-0.120***	-0.076***		-0.111***	-0.110***	-0.110***	-0.061***
		(0.023)	(0.023)	(0.023)	(0.020)		(0.020)	(0.020)	(0.020)	(0.017)
The student's mother is literate		0.085**	0.082**	0.082**	0.040		0.077*	0.074*	0.074*	0.032
		(0.029)	(0.029)	(0.028)	(0.027)		(0.031)	(0.031)	(0.031)	(0.028)
The student participates in commercial activities		0.026	0.027	0.027	-0.010		-0.007	-0.007	-0.007	-0.052
		(0.030)	(0.030)	(0.030)	(0.029)		(0.032)	(0.032)	(0.032)	(0.035)
Female teacher			0.191*	0.183+	0.135			0.112	0.117	0.088
			(0.092)	(0.095)	(0.083)			(0.087)	(0.091)	(0.077)
Female student x Female teacher			-0.046	-0.046	-0.016			-0.060	-0.060	-0.023
			(0.051)	(0.051)	(0.048)			(0.066)	(0.066)	(0.066)
Equipment level of the classroom			0.050	0.043	0.041			0.025	0.023	0.024
			(0.033)	(0.034)	(0.029)			(0.031)	(0.032)	(0.031)
Classroom-level average of the socioeconomic status of students' families			0.115	0.026	0.086			0.223*	0.198+	0.213*
			(0.118)	(0.133)	(0.111)			(0.101)	(0.114)	(0.094)
The number of students within the classroom is less than 40			0.066	0.066	0.049			0.046	0.045	0.016
			(0.058)	(0.058)	(0.054)			(0.059)	(0.059)	(0.056)
Experience of the principal (in years)				-0.007	-0.005				-0.005	-0.003
				(0.014)	(0.009)				(0.013)	(0.008)
School size (in hundreds of students)				-0.013	-0.009				0.019	0.023
				(0.044)	(0.031)				(0.034)	(0.024)
Equipment index of the school				0.079	0.036				0.082	0.044
				(0.081)	(0.067)				(0.082)	(0.067)
The school is located in a rural area				-0.115	-0.020				0.127	0.206+
				(0.200)	(0.145)				(0.160)	(0.113)
Standardized value of initial performance in French test score					0.289***					0.243***
					(0.021)					(0.023)
Standardized value of initial performance in mathematics test score					0.204***					0.301***
					(0.022)					(0.027)
Intercept	-0.019	-0.255**	-0.317**	-0.179	-0.059	-0.011	-0.321***	-0.346***	-0.420*	-0.278*
	(0.068)	(0.096)	(0.102)	(0.245)	(0.174)	(0.062)	(0.094)	(0.099)	(0.184)	(0.132)
lns1_1_1_cons	-0.420***	-0.422***	-0.449***	-0.454***	-0.866***	-0.540***	-0.540***	-0.569***	-0.580***	-1.110***
	(0.094)	(0.094)	(0.091)	(0.089)	(0.111)	(0.094)	(0.095)	(0.091)	(0.089)	(0.144)
lns2_1_1_cons	-0.741***	-0.742***	-0.735***	-0.740***	-0.846***	-0.756***	-0.751***	-0.763***	-0.763***	-0.789***
	(0.108)	(0.110)	(0.111)	(0.111)	(0.122)	(0.109)	(0.112)	(0.118)	(0.117)	(0.095)
lnsig_e _cons	-0.542***	-0.559***	-0.560***	-0.560***	-0.666***	-0.420***	-0.443***	-0.443***	-0.443***	-0.554***
	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)
Akaike information criterion	5,927.469	5,849.432	5,849.576	5,855.696	5,176.695	6,570.056	6,459.543	6,459.177	6,465.324	5,776.238
Observations	2,966	2,966	2,966	2,966	2,966	2,966	2,966	2,966	2,966	2966

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.6: *Estimated Models for Congo*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Grade dummy (1=Grade 5)		0.112+	0.164+	0.130	0.118		0.004	0.037	0.008	-0.008
		(0.065)	(0.084)	(0.083)	(0.072)		(0.062)	(0.078)	(0.079)	(0.071)
Age of the student (in years)		-0.018+	-0.018+	-0.017	-0.017+		0.011	0.012	0.013	0.011
		(0.010)	(0.011)	(0.011)	(0.009)		(0.011)	(0.011)	(0.011)	(0.009)
Socioeconomic status of the student's family		0.060***	0.050**	0.050***	0.039**		0.062***	0.053**	0.053**	0.042**
		(0.015)	(0.015)	(0.015)	(0.015)		(0.016)	(0.016)	(0.016)	(0.015)
The student gets help with schoolwork at home		0.013	0.013	0.013	0.003		0.028	0.028	0.028	0.019
		(0.023)	(0.023)	(0.023)	(0.023)		(0.025)	(0.026)	(0.026)	(0.025)
Number of times the student repeats a grade		-0.042**	-0.042**	-0.041**	-0.016		-0.046**	-0.046**	-0.046**	-0.019
		(0.015)	(0.015)	(0.015)	(0.013)		(0.014)	(0.014)	(0.014)	(0.013)
The student participates in field work		-0.024	-0.019	-0.017	-0.011		-0.022	-0.017	-0.016	-0.007
		(0.036)	(0.036)	(0.036)	(0.032)		(0.039)	(0.039)	(0.039)	(0.036)
The student uses the French language at home		0.106**	0.102**	0.100*	0.024		0.062	0.059	0.058	-0.016
		(0.040)	(0.040)	(0.040)	(0.031)		(0.038)	(0.038)	(0.038)	(0.033)
Female teacher			0.148+	0.099	0.096			0.058	0.016	0.006
			(0.083)	(0.082)	(0.073)			(0.081)	(0.085)	(0.074)
The teacher holds at least a Baccalaureate			0.042	0.054	0.034			-0.041	-0.033	-0.048
(Bac) degree			(0.068)	(0.066)	(0.061)			(0.071)	(0.069)	(0.065)
The teacher is a public servant[10]			-0.037	0.222	0.230			-0.177	0.009	0.091
			(0.163)	(0.202)	(0.187)			(0.148)	(0.152)	(0.145)
Contractual teacher			-0.091	0.160	0.195			-0.414**	-0.239	-0.099
			(0.242)	(0.270)	(0.286)			(0.152)	(0.154)	(0.183)
Community teacher			0.239	0.468*	0.458*			0.073	0.231	0.288+
			(0.167)	(0.197)	(0.182)			(0.156)	(0.156)	(0.157)
Experience of the teacher (in years)			0.009*	0.007+	0.006			0.006	0.005	0.004
			(0.004)	(0.004)	(0.004)			(0.005)	(0.005)	(0.005)
Equipment level of the classroom			0.011	-0.002	0.010			0.026	0.015	0.015
			(0.034)	(0.033)	(0.029)			(0.035)	(0.035)	(0.031)
Classroom-level average of the socioeconomic			0.398***	0.278**	0.214*			0.276**	0.165	0.123
status of students' families			(0.093)	(0.098)	(0.089)			(0.100)	(0.112)	(0.094)
Classroom-level standard deviation of the			-0.485**	-0.509***	-0.372**			-0.341*	-0.367*	-0.285*
socioeconomic status			(0.153)	(0.154)	(0.136)			(0.156)	(0.154)	(0.131)
The number of students within the			0.078	0.089	0.111			0.140+	0.155+	0.136+
classroom is less than 40			(0.084)	(0.084)	(0.076)			(0.083)	(0.086)	(0.070)
The principal is an official servant				-0.165	-0.090				-0.134	-0.105
				(0.163)	(0.128)				(0.136)	(0.108)
School size (in hundreds of students)				-0.019+	-0.019*				-0.008	-0.010
				(0.011)	(0.009)				(0.010)	(0.008)

Table C.6: *Continued*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Equipment index of the school				0.089	0.042				0.101	0.058
				(0.069)	(0.054)				(0.068)	(0.052)
The school is located in a rural area				-0.379*	-0.325**				-0.258+	-0.214+
				(0.155)	(0.119)				(0.152)	(0.121)
Standardized value of initial performance in French test score					0.292***					0.214***
					(0.021)					(0.021)
Standardized value of initial performance in mathematics test score					0.164***					0.255***
					(0.019)					(0.020)
Intercept	-0.030	0.089	0.092	0.325+	0.214	-0.032	-0.129	-0.005	0.142	0.081
	(0.064)	(0.107)	(0.190)	(0.187)	(0.161)	(0.060)	(0.101)	(0.179)	(0.193)	(0.164)
lns1_1_1_cons	-0.372***	-0.431***	-0.557***	-0.562***	-0.915***	-0.434***	-0.474***	-0.599***	-0.607***	-0.991***
	(0.078)	(0.080)	(0.085)	(0.082)	(0.101)	(0.085)	(0.087)	(0.092)	(0.090)	(0.114)
lns2_1_1_cons	-1.091***	-1.105***	-1.114***	-1.145***	-1.136***	-1.109***	-1.108***	-1.095***	-1.110***	-1.099***
	(0.074)	(0.075)	(0.086)	(0.087)	(0.089)	(0.078)	(0.078)	(0.087)	(0.089)	(0.082)
lnsig_e _cons	-0.479***	-0.485***	-0.485***	-0.485***	-0.576***	-0.408***	-0.411***	-0.411***	-0.411***	-0.496***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.030)	(0.026)	(0.026)	(0.026)	(0.026)	(0.028)
Akaike information criterion	7,229.537	7,187.571	7,175.424	7,174.436	6,520.305	7,674.293	7,655.934	7,649.879	7,652.293	7,032.135
Observations	3,472	3,472	3,472	3,472	3,472	3,472	3,472	3,472	3,472	3,472

$+\ p < 0.10$, $*\ p < 0.05$, $**\ p < 0.01$, $***\ p < 0.001$

[10] A teacher who is a public servant is one hired by the government for a career in education/for an extended period of time. A contractual teacher is one with a government contract who is hired for a limited period. A community teacher is hired by the local community.

Table C.7: *Estimated Models for Ivory Coast*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Grade dummy (1=Grade 5)		0.045	0.062	0.041	0.092+		-0.159*	-0.079	-0.096	-0.046
		(0.067)	(0.079)	(0.077)	(0.056)		(0.065)	(0.071)	(0.070)	(0.059)
Female student		-0.012	-0.025	-0.024	0.023		-0.100**	-0.136***	-0.135***	-0.057*
		(0.027)	(0.031)	(0.030)	(0.024)		(0.030)	(0.034)	(0.034)	(0.027)
Age of the student (in years)		0.004	0.005	0.005	-0.026***		0.056***	0.057***	0.057***	0.022*
		(0.012)	(0.012)	(0.012)	(0.008)		(0.013)	(0.013)	(0.013)	(0.009)
Socioeconomic status of the student's family		0.050**	0.018	0.019	-0.002		0.081***	0.052**	0.053**	0.035*
		(0.016)	(0.017)	(0.017)	(0.014)		(0.019)	(0.020)	(0.020)	(0.017)
The student gets help with schoolwork at home		0.050	0.055+	0.057+	0.030		0.001	0.004	0.006	-0.015
		(0.032)	(0.032)	(0.032)	(0.023)		(0.034)	(0.034)	(0.034)	(0.027)
Number of times the student repeats a grade		-0.108***	-0.106***	-0.105***	-0.025*		-0.077***	-0.075***	-0.075***	0.003
		(0.016)	(0.016)	(0.016)	(0.012)		(0.018)	(0.018)	(0.018)	(0.014)
The student's mother is literate		0.104***	0.103***	0.102***	0.019		0.076*	0.075*	0.073*	-0.012
		(0.027)	(0.027)	(0.027)	(0.020)		(0.032)	(0.032)	(0.032)	(0.026)
The student participates in commercial activities		-0.102**	-0.095**	-0.093**	-0.050*		-0.040	-0.032	-0.030	0.011
		(0.035)	(0.034)	(0.034)	(0.026)		(0.037)	(0.037)	(0.037)	(0.030)
The student participates in field work		-0.064+	-0.041	-0.028	-0.034		-0.037	-0.015	-0.002	-0.017
		(0.034)	(0.036)	(0.035)	(0.024)		(0.036)	(0.037)	(0.037)	(0.027)
The student uses the French language at home		0.226***	0.213***	0.208***	0.053*		0.103*	0.087*	0.082+	-0.060*
		(0.039)	(0.038)	(0.038)	(0.024)		(0.044)	(0.043)	(0.043)	(0.029)
Female teacher			-0.004	-0.031	-0.044			0.086	0.064	0.018
			(0.073)	(0.073)	(0.054)			(0.069)	(0.069)	(0.056)
Female student x Female teacher			0.029	0.029	0.010			0.095	0.095	0.072+
			(0.060)	(0.060)	(0.043)			(0.063)	(0.063)	(0.041)
Equipment level of the classroom			0.023	0.009	0.019			0.022	0.011	0.021
			(0.036)	(0.035)	(0.022)			(0.034)	(0.033)	(0.020)
Classroom-level average of the socioeconomic status of students' families			0.342***	0.151+	0.108*			0.243***	0.063	0.026
			(0.064)	(0.087)	(0.055)			(0.061)	(0.084)	(0.054)
Classroom-level standard deviation of the socioeconomic status			-0.284	-0.151	-0.047			-0.211	-0.096	-0.023
			(0.188)	(0.168)	(0.102)			(0.184)	(0.166)	(0.100)
The number of students within the classroom is less than 40			-0.110	-0.082	-0.006			-0.113	-0.096	-0.018
			(0.069)	(0.063)	(0.047)			(0.076)	(0.072)	(0.053)
The teacher has a pedagogical diploma			0.036	0.058	0.044			-0.055	-0.042	-0.010
			(0.085)	(0.081)	(0.064)			(0.080)	(0.077)	(0.059)
Experience of the principal (in years)				-0.006	-0.009*				-0.004	-0.007+
				(0.006)	(0.004)				(0.007)	(0.004)
School size (in hundreds of students)				0.004	-0.027				-0.020	-0.050*
				(0.045)	(0.027)				(0.041)	(0.025)

Table C.7: *Continued*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Equipment index of the school				0.202**	0.079+				0.157**	0.048
				(0.062)	(0.041)				(0.056)	(0.036)
The school is located in a rural area				-0.086	0.032				-0.176	-0.062
				(0.127)	(0.088)				(0.127)	(0.079)
Standardized value of initial performance					0.522***					0.350***
in French test score					(0.018)					(0.017)
Standardized value of initial performance					0.208***					0.373***
in mathematics test score					(0.015)					(0.018)
Intercept	-0.011	-0.059	0.069	0.056	0.336*	-0.009	-0.407***	-0.301*	-0.190	0.114
	(0.053)	(0.107)	(0.155)	(0.230)	(0.151)	(0.048)	(0.113)	(0.152)	(0.225)	(0.145)
lns1_1_1_cons	-0.589***	-0.757***	-0.979***	-1.008***	-1.748***	-0.708***	-0.805***	-0.927***	-0.959***	-1.737***
	(0.075)	(0.087)	(0.104)	(0.110)	(0.174)	(0.074)	(0.083)	(0.096)	(0.106)	(0.187)
lns2_1_1_cons	-1.086***	-1.096***	-1.055***	-1.094***	-1.340***	-1.125***	-1.130***	-1.125***	-1.150***	-1.281***
	(0.074)	(0.072)	(0.071)	(0.073)	(0.079)	(0.083)	(0.083)	(0.085)	(0.086)	(0.085)
lnsig_e _cons	-0.275***	-0.288***	-0.288***	-0.289***	-0.593***	-0.212***	-0.221***	-0.221***	-0.222***	-0.467***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.017)	(0.019)	(0.019)	(0.019)	(0.019)	(0.015)
Akaike information criterion	9,349.499	9,236.644	9,213.807	9,204.818	6,833.524	9,766.081	9,701.840	9,686.809	9,681.579	7,763.729
Observations	3,825	3,825	3,825	3,825	3,825	3,825	3,825	3,825	3,825	3,825

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.8: *Estimated Models for Senegal*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Grade dummy (1=Grade 5)		0.015	0.008	0.014	0.078		-0.126+	-0.140+	-0.141+	-0.073
		(0.070)	(0.085)	(0.085)	(0.062)		(0.069)	(0.082)	(0.082)	(0.063)
Female student		-0.065*	-0.098**	-0.098**	-0.017		-0.124***	-0.170***	-0.170***	-0.085**
		(0.031)	(0.036)	(0.036)	(0.026)		(0.031)	(0.034)	(0.035)	(0.028)
Age of the student		0.012	0.011	0.012	-0.028**		0.048***	0.046**	0.048***	0.007
		(0.013)	(0.013)	(0.013)	(0.010)		(0.014)	(0.014)	(0.014)	(0.012)
Socioeconomic status of the student's family		0.090***	0.064**	0.064**	0.013		0.073**	0.039+	0.039+	-0.011
		(0.023)	(0.024)	(0.024)	(0.018)		(0.022)	(0.022)	(0.022)	(0.016)
The student gets help with schoolwork at home		0.076+	0.072+	0.072+	0.053+		0.059	0.053	0.054	0.035
		(0.042)	(0.042)	(0.042)	(0.030)		(0.041)	(0.041)	(0.041)	(0.030)
Number of times the student repeats a grade		-0.160***	-0.156***	-0.156***	-0.088***		-0.119***	-0.115***	-0.115***	-0.049**
		(0.023)	(0.023)	(0.023)	(0.019)		(0.023)	(0.024)	(0.024)	(0.019)
The student's mother is literate		0.036	0.035	0.036	-0.008		-0.013	-0.014	-0.013	-0.050+
		(0.031)	(0.031)	(0.031)	(0.025)		(0.032)	(0.032)	(0.032)	(0.026)
The student participates in commercial activities		-0.029	-0.029	-0.028	-0.013		-0.017	-0.019	-0.018	-0.005
		(0.039)	(0.039)	(0.039)	(0.032)		(0.047)	(0.046)	(0.046)	(0.034)
The student participates in field work		-0.117**	-0.096*	-0.098*	-0.055+		-0.120**	-0.090*	-0.095*	-0.053
		(0.039)	(0.041)	(0.040)	(0.030)		(0.038)	(0.040)	(0.040)	(0.033)
Female teacher			-0.076	-0.033	-0.117+			-0.082	-0.043	-0.126*
			(0.096)	(0.094)	(0.068)			(0.092)	(0.091)	(0.064)
Female student x Female teacher			0.104+	0.103+	0.055			0.145*	0.144*	0.100*
			(0.061)	(0.061)	(0.046)			(0.060)	(0.060)	(0.047)
Equipment level of the classroom			0.037	0.039	0.019			0.059+	0.060+	0.026
			(0.039)	(0.037)	(0.035)			(0.034)	(0.033)	(0.030)
Classroom-level average of the socioeconomic status of students' families			0.301***	0.247**	0.133*			0.307***	0.285***	0.158**
			(0.062)	(0.080)	(0.052)			(0.059)	(0.080)	(0.054)
Classroom-level standard deviation of the socioeconomic status			-0.148	-0.077	-0.090			-0.074	0.001	-0.006
			(0.139)	(0.135)	(0.098)			(0.152)	(0.151)	(0.116)
The number of students within the classroom is less than 40			0.208*	0.165	0.089			0.173*	0.105	-0.003
			(0.095)	(0.102)	(0.075)			(0.087)	(0.097)	(0.074)
The teacher has a pedagogical diploma			-0.016	0.026	0.029			0.040	0.101	0.111+
			(0.088)	(0.090)	(0.062)			(0.081)	(0.082)	(0.057)
Experience of the principal (in years)				0.006	0.001				0.001	-0.005
				(0.006)	(0.004)				(0.005)	(0.004)
Principal with at least BFEM[11]				-0.017	-0.023				-0.024	-0.031
				(0.088)	(0.057)				(0.082)	(0.056)
The principal is an official servant				-0.197	-0.070				-0.194	-0.126
				(0.138)	(0.085)				(0.131)	(0.080)
School size (in hundreds of students)				-0.023	-0.010				-0.020	-0.010
				(0.014)	(0.009)				(0.014)	(0.012)

Table C.8: *Continued*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Equipment index of the school				0.172***	0.068+				0.145**	0.051
				(0.050)	(0.036)				(0.047)	(0.035)
The school is located in a rural area				0.220+	0.190**				0.262*	0.229**
				(0.121)	(0.070)				(0.129)	(0.077)
Standardized value of initial performance in French test score					0.463***					0.342***
					(0.017)					(0.019)
Standardized value of initial performance in mathematics test score					0.233***					0.357***
					(0.019)					(0.023)
Intercept	0.011	0.012	0.085	0.144	0.391*	0.007	-0.245+	-0.242	-0.165	0.159
	(0.055)	(0.137)	(0.184)	(0.243)	(0.154)	(0.051)	(0.143)	(0.189)	(0.235)	(0.162)
lns1_1_1_cons	-0.591***	-0.709***	-0.913***	-1.010***	-1.683***	-0.713***	-0.811***	-1.019***	-1.128***	-1.842***
	(0.089)	(0.097)	(0.117)	(0.133)	(0.221)	(0.093)	(0.107)	(0.130)	(0.139)	(0.259)
lns2_1_1_cons	-0.918***	-0.922***	-0.886***	-0.897***	-1.114***	-0.924***	-0.910***	-0.897***	-0.908***	-1.084***
	(0.086)	(0.085)	(0.087)	(0.086)	(0.093)	(0.090)	(0.089)	(0.086)	(0.088)	(0.068)
lnsig_e _cons	-0.297***	-0.309***	-0.310***	-0.310***	-0.567***	-0.241***	-0.250***	-0.251***	-0.251***	-0.471***
	(0.020)	(0.019)	(0.019)	(0.019)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Akaike information criterion	8,639.702	8,552.106	8,535.296	8,528.674	6,659.842	8,985.231	8,929.778	8,906.029	8,899.167	7,300.634
Observations	3,575	3,575	3,575	3,575	3,575	3,575	3,575	3,575	3,575	3,575

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[11] BFEM is Brevet de Fin d'Etudes Moyennes. It is a secondary school certificate obtained after 10 years of schooling.

Table C.9: *Estimated Models for Togo*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Grade dummy (1=Grade 5)		0.060	0.019	0.027	0.057		-0.083	-0.126*	-0.122*	-0.076
		(0.060)	(0.058)	(0.058)	(0.050)		(0.054)	(0.055)	(0.054)	(0.050)
Female student		-0.095***	-0.102**	-0.103**	0.014		-0.149***	-0.152***	-0.152***	-0.032
		(0.025)	(0.034)	(0.034)	(0.027)		(0.027)	(0.037)	(0.037)	(0.028)
Age of the student		0.003	0.007	0.007	-0.009		0.040***	0.043***	0.043***	0.024**
		(0.009)	(0.009)	(0.009)	(0.008)		(0.009)	(0.009)	(0.009)	(0.008)
Socioeconomic status of the student's family		0.137***	0.090***	0.090***	0.022		0.106***	0.070***	0.070***	0.011
		(0.020)	(0.020)	(0.020)	(0.015)		(0.019)	(0.020)	(0.020)	(0.016)
Number of times the student repeats a grade		-0.179***	-0.177***	-0.176***	-0.112***		-0.141***	-0.138***	-0.138***	-0.071***
		(0.021)	(0.021)	(0.021)	(0.015)		(0.020)	(0.020)	(0.020)	(0.015)
The student participates in field work		-0.064*	-0.036	-0.035	-0.052*		-0.044	-0.023	-0.022	-0.040
		(0.031)	(0.031)	(0.031)	(0.025)		(0.034)	(0.034)	(0.034)	(0.028)
Female teacher			-0.012	-0.031	0.003			-0.021	-0.031	0.009
			(0.058)	(0.060)	(0.050)			(0.060)	(0.059)	(0.051)
Female student x Female teacher			0.010	0.011	-0.031			0.007	0.007	-0.028
			(0.049)	(0.049)	(0.037)			(0.055)	(0.055)	(0.043)
Experience of the teacher (in years)			-0.012***	-0.012***	-0.010***			-0.005+	-0.006+	-0.003
			(0.003)	(0.003)	(0.003)			(0.003)	(0.003)	(0.002)
Equipment level of the classroom			0.008	-0.013	-0.014			-0.020	-0.036	-0.040
			(0.030)	(0.033)	(0.027)			(0.033)	(0.035)	(0.030)
Classroom-level average of the socioeconomic status of students' families			0.419***	0.380***	0.138+			0.336***	0.309***	0.068
			(0.076)	(0.089)	(0.072)			(0.074)	(0.087)	(0.069)
Classroom-level standard deviation of the socioeconomic status			0.032	0.032	0.053			-0.113	-0.108	-0.055
			(0.147)	(0.148)	(0.126)			(0.139)	(0.138)	(0.130)
The number of students within the classroom is less than 40			0.097+	0.062	0.077			0.141*	0.128*	0.129*
			(0.051)	(0.056)	(0.047)			(0.061)	(0.064)	(0.056)
The teacher has at least the BEPC			0.002	0.011	-0.026			-0.080	-0.070	-0.095
			(0.052)	(0.057)	(0.046)			(0.062)	(0.065)	(0.058)
Experience of the principal (in years)				0.002	-0.004				0.003	-0.003
				(0.005)	(0.003)				(0.006)	(0.003)
The principal has at least the BEPC				-0.056	0.024				-0.093	-0.007
				(0.071)	(0.054)				(0.081)	(0.063)
School size (in hundreds of students)				-0.072	-0.039				-0.019	0.001
				(0.076)	(0.055)				(0.097)	(0.076)

Table C.9: *Continued*

	French					Mathematics				
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4
Equipment index of the school				0.063	0.034				0.063	0.035
				(0.045)	(0.032)				(0.043)	(0.034)
The school is located in a rural area				0.005	-0.110+				0.050	-0.058
				(0.097)	(0.066)				(0.097)	(0.068)
Standardized value of initial performance in French test score					0.402***					0.269***
					(0.019)					(0.017)
Standardized value of initial performance in mathematics test score					0.273***					0.407***
					(0.017)					(0.018)
Intercept	-0.010	0.140	0.186	0.301	0.372**	-0.006	-0.149+	-0.056	-0.032	0.038
	(0.047)	(0.086)	(0.127)	(0.192)	(0.141)	(0.043)	(0.084)	(0.128)	(0.208)	(0.165)
lns1_1_1_cons	-0.632***	-0.880***	-1.302***	-1.301***	-2.009***	-0.726***	-0.829***	-0.968***	-0.980***	-1.494***
	(0.095)	(0.112)	(0.137)	(0.140)	(0.451)	(0.072)	(0.080)	(0.090)	(0.093)	(0.185)
lns2_1_1_cons	-0.973***	-0.960***	-0.973***	-0.984***	-1.074***	-1.062***	-1.062***	-1.077***	-1.081***	-1.020***
	(0.114)	(0.109)	(0.108)	(0.110)	(0.136)	(0.081)	(0.078)	(0.075)	(0.076)	(0.080)
lnsig_e _cons	-0.279***	-0.294***	-0.295***	-0.295***	-0.541***	-0.223***	-0.235***	-0.236***	-0.236***	-0.454***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.022)	(0.015)	(0.014)	(0.014)	(0.014)	(0.017)
Akaike information criterion	11,092.600	10,921.483	10,844.138	10,850.116	8,640.778	11,517.655	11,398.819	11,372.729	11,378.952	9,456.296
Observations	4,540	4,540	4,540	4,540	4,540	4,540	4,540	4,540	4,540	4,540

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.10: *Determinants of Academic Growth between Grades 2 and 5, French Test Score*

	Cameroon	Chad	Congo	Ivory Coast	Senegal	Togo
Female principal	0.329+					
	(0.190)					
Experience of the principal (in years)	-0.002	0.007	-0.001	0.000	-0.020**	-0.005
	(0.009)	(0.013)	(0.007)	(0.008)	(0.007)	(0.005)
The principal holds at least the BEPC or an equivalent diploma	0.009				0.044	0.075
	(0.129)				(0.107)	(0.099)
The principal is an official servant	-0.118		0.096		0.054	
	(0.157)		(0.135)		(0.156)	
Change in socioeconomic status between grades 2 and 5	-0.033	-0.011	0.054	-0.080	0.172	0.012
	(0.195)	(0.244)	(0.129)	(0.133)	(0.165)	(0.141)
Average socioeconomic status in the school		0.159	0.354***	0.028	0.034	0.165*
		(0.151)	(0.104)	(0.087)	(0.106)	(0.084)
School size (in hundreds of students)	0.023	0.000	-0.000	-0.000	0.002	0.001
	(0.037)	(0.000)	(0.000)	(0.001)	(0.019)	(0.001)
There is a management committee in the school	-0.279+		-0.098			
	(0.164)		(0.107)			
Equipment index of the school	-0.024	-0.112	-0.101+	0.036	0.090	-0.058
	(0.069)	(0.077)	(0.059)	(0.073)	(0.072)	(0.048)
The school is located in a rural area	0.085	0.206	0.261+	0.091	0.165	-0.120
	(0.161)	(0.190)	(0.154)	(0.145)	(0.176)	(0.151)
Equipment index of the locality where the school is located	0.033	-0.014	-0.068	0.050	-0.070	0.023
	(0.077)	(0.075)	(0.064)	(0.061)	(0.070)	(0.062)

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.11: *Determinants of Academic Growth between Grades 2 and 5, Mathematics Test Score*

	Cameroon	Chad	Congo	Ivory Coast	Senegal	Togo
Female principal	0.191 (0.242)					
Experience of the principal (in years)	0.004 (0.010)	0.006 (0.012)	0.004 (0.007)	-0.002 (0.008)	-0.014* (0.006)	0.006 (0.006)
The principal holds at least the BEPC or an equivalent diploma	0.019 (0.148)				0.068 (0.103)	0.084 (0.094)
The principal is an official servant	-0.123 (0.166)		-0.054 (0.148)		0.126 (0.174)	
Change in socioeconomic status between grades 2 and 5	0.168 (0.173)	0.397* (0.189)	-0.026 (0.146)	-0.011 (0.131)	0.294+ (0.178)	0.035 (0.132)
Average socioeconomic status in the school		0.087 (0.160)	0.162 (0.126)	-0.033 (0.094)	0.054 (0.124)	0.157+ (0.084)
School size (in hundreds of students)	0.002 (0.038)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.001)	-0.022 (0.015)	-0.001 (0.001)
There is a management committee in the school	0.169 (0.185)		0.007 (0.105)			
Equipment index of the school	-0.091 (0.065)	0.024 (0.094)	0.000 (0.064)	-0.102 (0.069)	0.104 (0.090)	-0.057 (0.047)
The school is located in a rural area	0.123 (0.134)	0.407** (0.155)	0.128 (0.151)	-0.122 (0.154)	0.121 (0.169)	0.081 (0.139)
Equipment index of the locality where the school is located	0.057 (0.079)	0.019 (0.074)	-0.056 (0.063)	0.023 (0.070)	-0.048 (0.068)	-0.054 (0.057)

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$