EDUCATIONAL EFFECTIVENESS AND INEQUALITIES IN CHILE:
A Multilevel Accelerated Longitudinal Study of Primary School Children's Achievement Trajectories

Lorena Constanza Ortega Ferrand

Department of Education

University of Oxford

St Anne's College

Supervisors

Dr Lars-Erik Malmberg

Professor Pam Sammons

Thesis submitted to the University of Oxford in fulfilment of the requirements for the degree of Doctor of Philosophy in Education

Hilary Term, 2016

ABSTRACT

Investigating the impact of schools and teachers on student achievement has become an international trend, and Educational Effectiveness Research (EER) has generally found these effects to be modest in size. The field has undergone significant methodological advances and developed new methods for estimating educational effects, favouring the study of students' growth trajectories using a multilevel longitudinal approach. This method is able to demonstrate more sizeable school and teacher effects. However, most educational longitudinal research comes from post-industrialised countries. Thus, it is still unclear whether the claims produced by this body of knowledge are pertinent to emerging economies.

The present study investigates educational effects of both schools and teachers on primary students' achievement trajectories in Chile: a context of particular interest, given its socially stratified and segregated schooling system and its unregulated and diverse teacher labour force. Several properties of school and teacher effects, such as magnitude, consistency, predictors, cumulativeness and differential effects across student groups within schools are investigated using a series of multilevel growth models.

By means of linking several sources of secondary data, of which some have not been used for research purposes before, rich longitudinal data on student achievement in language (Spanish) and mathematics were obtained. The resulting sample features an accelerated longitudinal design comprising of participants in 4 overlapping cohorts, together spanning Grades 3 to 8 (N = 19,704 students, and 851 language and 812 mathematics teachers, in 156 schools) and incorporates a wide range of schools, teachers and students. The quality of the data allows for the modelling of school and teacher effects on student achievement growth over time, which represents a clear improvement when compared to previous measures of educational effectiveness developed in the Chilean and Latin-American context, which cover two time points at most. Furthermore, the study is the first in the region that annually matches students with their teachers, and models the relationships between students and their successive classroom settings.

This study's main findings on student achievement trajectories indicate non-linear upward growth on student achievement for both language and mathematics in primary school. In addition, individual students differ substantially in both their achievement status and their rate of development over time. In language, a gender gap favouring girls that remains stable across primary school was found.

In mathematics, in turn, the gender gap reverses in favour of boys and increases from 3rd to 8th grade. An achievement gap between high- and low-socioeconomic status (SES) students is also present from 3rd grade, and remains fairly constant over the course of the primary school years.

School effects on students' growth trajectories were found to be sizeable (in fact, larger than those found in previous studies using similar model specifications and outcomes in the other national contexts) and moderately consistent across the two subjects. Evidence of compositional effects was found, as school achievement mean predicted achievement status on both subjects. Also, in language, the school's SES composition was found to have effects on achievement outcomes over and above the individual student's SES, supporting the double jeopardy hypothesis. The results also show that school sector (i.e., public vs. private school) differences on student achievement are largely due to differences in student intake and not to differences in school effectiveness. In both subjects, schools were found to be differentially effective across students from different socioeconomic status. In language, schools also showed differential effects associated with student gender. In addition, it was found that teacher effects in the primary school level are large, exceeding school effects, and not highly consistent across subjects. Finally, the contribution of teachers to student achievement growth was found to accumulate over time.

The findings from this study add to the evidence that longitudinal studies examining student growth are more likely to demonstrate educational effects of larger magnitude than studies using covariate-adjustment and gain scores models over just two time points. This confirms that both school and teacher effects are important in shaping achievement growth. The findings also demonstrate that school and teacher contributions have stronger effects on student achievement growth than on achievement status.

This study addresses three important gaps in the literature. Firstly, it explores educational effects in the context of an emerging economy, using appropriate EER models in terms of measures and specifications. Secondly, it contributes further evidence on the properties of school and teacher effects on student achievement growth. Thirdly, it advances the field methodologically by demonstrating the combined use of accelerated longitudinal designs, growth curve approaches, and cross-classified and multiple membership models.

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to my supervisors Dr Lars-Erik Malmberg and Professor Pam Sammons for their invaluable support and guidance. Their sound advice, ideas, knowledge and teaching of the literature and state-of-the-art statistical techniques, as well as their care and support, contributed immensely to the completion of my thesis and made my doctoral studies a stimulating, enjoyable and productive experience.

My sincere gratitude goes to the Chilean scholarship programme BECAS CHILE, as without their financial support this research project would not have been possible. I would also like to thank St Anne's College, University of Oxford, for granting funds for conference and workshop attendance.

I am heartily thankful to all the colleagues and friends that have supported me in various invaluable ways. I would like to especially thank Daniel Caro, Diego Carrasco, Teresa Florez, Gianna Hessel and Tatiana Rodriguez for the helpful advice and emotional support they provided.

Finally, I cannot thank Patricio and the rest of my family in Chile enough for all the love, patience, understanding and encouragement they have provided me during my doctoral studies and throughout my life. They have always been there for me and I am deeply grateful for that. This thesis is dedicated to my son Pablo, my greatest motivation in studies, work and life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF APPENDICES

LIST OF ABREVIATIONS

BERA	*British Educational Research Association*
CNE	*Consejo Nacional de Educación [Chilean National Council of Education]*
CUREC	*Central University Research Ethics Committee*
CVAM	*Contextualised Value-Added Models*
DREC	*Departmental Research Ethics Committee*
EER	*Educational Effectiveness Research*
ICC	*Intraclass Correlation Coefficients*
IEA	*International Association for the Evaluation of Educational Achievement*
IRT	*Item Response Theory*
ITT	*Initial Teacher Training*
MAR	*Missing at Random*
MCAR	*Missing Completely at Random*
MCMC	*Markov chain Monte Carlo*
MI	*Multiple Imputation*
MIDE-UC	*Assessment Centre at the Pontificia Universidad Católica de Chile*
MINEDUC	*Ministerio de Educación [Ministry of Education]*
MLwiN	*Statistical package for fitting multilevel models*
NMAR	*Not Missing at Random*
OECD	*Organisation for Economic Co-operation and Development*
PISA	*Programme for International Student Assessment*
PSU	*Prueba de Selección Universitaria [University Selection Test]*
SEPA	*Sistema de Evaluación de Progreso en el Aprendizaje [Learning Progress Evaluation System]*
SES	*Socio-Economic Status*
SIGE	*Sistema de Información General de Estudiantes [General System of Student Information]*
SIMCE	*Sistema Nacional de Evaluación de Calidad Educativa [Educational Quality Measuring System]*
SPSS	*Statistical Package for the Social Sciences*
TALIS	*Teacher and Learning International Survey*
TEDS-M	*Teacher Education and Development Study in Mathematics*
TIMSS	*Trends in International Mathematics and Science Study*
UNESCO	*United Nations Educational, Scientific and Cultural Organization*
VAM	*Value-Added Models*
VPC	*Variance Partition Coefficient*

CHAPTER 1: INTRODUCTION

This introductory chapter starts with the rationale for undertaking the research that is documented in this thesis. It then provides further information on the research context and concludes with an outline of the thesis in order to orientate the reader.

1.1 RATIONALE

In Latin America, several countries are experiencing economic growth and poverty reduction, but large socio-economic inequalities still persist and represent a great challenge for long-term sustainable growth. In the recent Lima Declaration (UNESCO, 2014), the governments from the countries of Latin America and the Caribbean set an urgent priority: to work on educational quality, inclusivity and equity in order to guarantee the right to education and contribute to reducing inequality and poverty in the region. This call for action is in line with the evidence from previous large-scale comparative studies, showing highly heterogeneous student learning results, and highlighting the sizable disparities and inequities present both among and within countries (Inter-American Development Bank, 2014; Murillo and Roman, 2011; UNESCO, 2008b; 2015a; Willms and Somer, 2001).

Whilst most countries in the region have achieved universal primary education and are witnessing a rapid expansion of pre-primary, secondary and tertiary education, learning achievement levels remain low by international standards and there is little evidence that attainment gaps are reducing (Cox, 2010; Mullis et al., 2012; OECD, 2013a; 2014a). In fact, the available evidence points to persistent and, in some countries, even increasing attainment gaps related to socio-economic background, gender, and schools in rural and urban areas (OECD/CAF/ECLAC, 2014). Indeed, socio-economic background and social environments are key markers of performance in Latin America, with the socio-economic status of the student and the school accounting for around 30% of the performance variation of secondary-school students in the region (OECD, 2013a).

The education systems of developing countries, such as Chile, have to deal with high levels of deprivation and multiple inequalities. Chile's inequality is reflected in the Gini coefficient[1] of 50.5 (World Bank, 2015) and a reduction of 19.3% in its Human Development Index (HDI) in 2014 due to inequality (United Nations Development Programme, 2015). Large HDI differences by region and between localities in the country have also been found, making it plausible to hypothesise the existence of significant differences in pupils' achievement according to the geographical location of their school.

Evidence from the UNESCO's Third Regional Comparative and Explanatory Study, (TERCE)[2] gives insights into the learning achievement levels of representative samples of Latin-American and Caribbean students in 15 countries. The study shows that, while Chile is among the highest-performing countries in all of the subjects and levels assessed, it is also one of the systems with the highest within-country variability in outcomes in the region (UNESCO, 2015c).

The substantive variations in children's access to educational resources and opportunities found in Chile affect children's skill formation and later opportunities in life, making equity in the distribution of learning a pending and urgent issue. Internationally, evidence shows that the highest-performing countries are those that allocate educational resources more equitably among socio-economically advantaged and disadvantaged schools (OECD, 2013). In order to foster further inclusive growth in the country, school-performance improvements must be accompanied by greater inclusion.

Furthermore, there is plenty of room for improvement in terms of overall levels of achievement, as Chilean students perform consistently below average in mathematics, reading and science in different international assessment studies such as the OECD

[1] The Gini index is a measure of inequality that ranges from 0 to 100, where 0 indicates complete equality and 100 indicates complete inequality. For a more thorough explanation, see The World Bank (2015).
[2] TERCE assessed in 2013 the educational attainment of pupils in third and sixth grade primary school in mathematics, reading, writing and natural sciences. The 15 participating countries were Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru and Uruguay.

Programme for International Student Assessment (PISA)[3] and the IEA Trends in International Mathematics and Science Study (TIMSS)[4] (Mullis et al., 2012; OECD, 2015b). Moreover, in Chile, the strength of the relationship between student performance and socio-economic status—an indication of inequality in education outcomes—is above the OECD average[5] and one of the strongest in Latin America (OECD, 2013a; UNESCO, 2015b). Indeed, Chile retains some of the starkest education disparities in the region, with a marked socioeconomic stratification of educational attainment (Cox, 2010; Mizala and Torche, 2012).

In this context, where social background is a strong predictor of students' school destination and attainment, it is relevant to investigate whether schools and teachers can make a difference, and identify the particular aspects of schools and teachers that hinder student learning progress, and those that "add value" to student performance beyond the levels expected based on student characteristics.

However, an important constraint to the analysis of educational effects in the region is the scarcity of longitudinal data on student achievement. In Chile, the existing educational effectiveness studies have all used cross-sectional data or, at most, two time points longitudinal data. The most commonly used data sources are the System for Measuring Educational Quality (SIMCE), administered by the Ministry of Education, and the cross-sectional international studies in which the country has participated, such as PISA, TIMSS, and the different waves of the UNESCO comparative studies (i.e., PERCE,

---

[3] PISA assessed in 2012 the educational attainment of 15-year-olds in mathematics, reading and sciences. The 63 participant countries were Albania, Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Colombia, Costa Rica, Croatia, Cyprus, Chinese Taipei, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong-China, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Shanghai (China), Jordan, Kazakhstan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Macao-China, Thailand, Malaysia, Mexico, Republic of Montenegro, Netherlands, New Zealand, Norway, Peru, Poland, Portugal, Qatar, Romania, Russian Federation, Republic of Serbia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Tunisia, Turkey, UAE, United Kingdom, United States, Uruguay and Vietnam.

[4] TIMSS assessed in 2011 the educational attainment of 4th- and 8th-graders in mathematics and science knowledge. The 63 participant countries and education systems were Armenia, Australia, Austria, Azerbaijan, Bahrain, Belgium (Flemish), Chile, Chinese Taipei, Croatia, Czech Republic, Denmark, England, Finland, Georgia, Germany, Ghana, Hong Kong SAR, Hungary, Indonesia, Iran, Islamic, Rep. of, Ireland, Rep. of, Israel, Italy, Japan, Jordan, Kazakhstan, Korea, Kuwait, Lebanon, Lithuania, Macedonia, Rep. of, Malaysia, Malta, Morocco, Netherlands, New Zealand, Northern Ireland, Norway, Oman, Palestinian Nat'l Auth., Poland, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Serbia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Syrian Arab Republic, Thailand, Tunisia, Turkey, Rep. of, Ukraine, United Arab Emirates, United States and Yemen.

[5] In Chile, 23.1% of the variation in performance is described by the variation in socio-economic status, whereas the OECD average is 14.8%.

SERCE and TERCE). As will be explained in the following chapters, these studies have not been able to appropriately disentangle the contribution of schools and teachers on student outcomes from the variance introduced by measurement error, and their results are likely to be biased, as cross-sectional and short longitudinal designs tend to underestimate educational effects (McCaffrey et al., 2003; Rowan, Correnti and Miller, 2002).

Investigating teacher effects in Chile is particularly relevant as teaching has become an unregulated profession, and the teacher labour force is currently highly diverse. Indeed, Chilean teachers are likely to differ considerably in key aspects, such as their initial training, disciplinary knowledge, teaching experience and work conditions. However, research on the impact of teachers on student outcomes in Chile is scarce, in part due to the difficulty of linking student and teacher data. The Ministry of Education collects data on teachers (e.g., demographics and professional training) and students' academic achievement separately. Matching students with their teachers (and schools) requires the use of unique identifiers across official databases. This capacity, only recently developed in Chile, is essential to investigate the associations between teachers and their students' attainment. Furthermore, while students within any particular school in Chile are usually taught by varied combinations of teachers over time, previous teacher effectiveness studies (e.g., Alvarado et al., 2012; Microdatos - University of Chile, 2006) have not been able to model the relationships between students and their successive classroom settings.

This thesis analyses in depth school and teacher effects on student achievement trajectories in Chile, and investigates educational inequalities by revealing the student, teacher and school factors associated with achievement growth. To this goal, the study uses linked longitudinal secondary data and state-of-the-art statistical tools and methods.

This research is relevant not only because it can advance current understanding of educational effectiveness in Chile, but also because it will examine equity aspects of the system. As explained above, the study of education inequalities is particularly relevant in Chile and Latin America where, despite universal access to primary education, rapid expansion of pre-primary education, and economic growth achieved in the past couple of decades, large differences between socio-economically advantaged and vulnerable groups persist and represent a great challenge for long-term sustainable growth (Inter-American Development Bank, 2014; OECD/CAF/ECLAC, 2014; UNESCO, 2015a). An adequate understanding of these relationships is expected to inform the debate concerned with narrowing the significant achievement gaps in the country, and help policymakers to improve decisions related to resourcing schools and supporting teachers.

Above, the motivation for conducting this research was introduced and its relevance pointed out. The following section presents the context of the study, namely, the Chilean education system.

1.2 RESEARCH CONTEXT

1.2.1 OVERVIEW OF THE CHILEAN EDUCATION SYSTEM

The specific context in which this study is situated is Chile. This section touches on relevant aspects of this educational system, such as the structure of the school system, the provision of initial teacher training and the composition of the teacher labour force. Using data from large-scale assessment programmes, such as PISA, TIMSS, the OECD Teaching and Learning International Survey (TALIS)[6] and the IEA Teacher Education and Development Study in Mathematics (TEDS-M)[7], the main features of the Chilean system are described from a comparative point of view.

1.2.1.1 STRUCTURE OF THE SCHOOL SYSTEM

The school system in Chile is organised in three sequential levels: pre-primary education (for children up to 5 years old), primary education (divided into 8 years with typical ages

[6] TALIS surveyed teachers and principals in schools in 2013 and focused on studying their working conditions and learning environments. The 34 participating countries were Australia, Belgium (Flanders), Brazil, Bulgaria, Canada (Alberta), Chile, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Iceland, Israel, Italy, Japan, Korea, Latvia, Malaysia, Mexico, Netherlands, Norway, Poland, Portugal, Serbia, Singapore, Slovak Republic, Romania, Spain, Sweden, United Arab Emirates, United Kingdom (England) and United States.

[7] TEDS-M assessed primary and secondary mathematics teachers in 2007-2008 in mathematics and pedagogy content knowledge. The 17 participating countries were Botswana, Canada, Chile, Chinese Taipei, Georgia, Germany, Malaysia, Norway, Oman, Philippines, Poland, Russian Federation, Singapore, Spain, Switzerland (German-speaking cantons), Thailand and United States.

6 to 13) and secondary education (divided into 4 years with typical ages 14 to 17) (see Figure 1). As of 2003, both primary and secondary education are mandatory for children up to 18 years old (12 years of compulsory schooling). By contrast, pre-primary education is not mandatory but it is free of charge. Enrolment in pre-primary programmes has increased over the years and is among the highest in Latin America, reaching 60% of 4-to-6 year-old children in 2013 (UNESCO, 2015b), but still remains below the OECD average (OECD, 2015a).

FIGURE 1: THE CHILEAN SCHOOL SYSTEM

<table>
<tr><th>Age</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th></tr>
<tr><th>Year</th><td></td><td></td><td></td><td>1st</td><td>2nd</td><td>3rd</td><td>4th</td><td>5th</td><td>6th</td><td>7th</td><td>8th</td><td>9th</td><td>10th</td><td>11th</td><td>12th</td></tr>
<tr><th rowspan="3">Level/strand</th><td colspan="3" rowspan="3">Pre-primary education</td><td colspan="8">Primary education</td><td colspan="4">Secondary education</td></tr>
<tr><td colspan="4" rowspan="2">First cycle</td><td colspan="4" rowspan="2">Second cycle</td><td colspan="2" rowspan="2">General (with both scientific-humanistic and technical-professional studies)</td><td colspan="2">Scientific-humanistic studies</td></tr>
<tr><td colspan="2">Technical professional studies</td></tr>
</table>

In Chile, governance of the school system is shared between central and local authorities. The Ministry of Education sets the central framework and the policy agenda, providing schools with a high level of autonomy over resource allocation, curriculum and assessment. Education is delivered by municipalities, and by a high proportion of privately managed educational institutions that receive public subsidies. The mixed system of private and public provision consists of three types of schools:

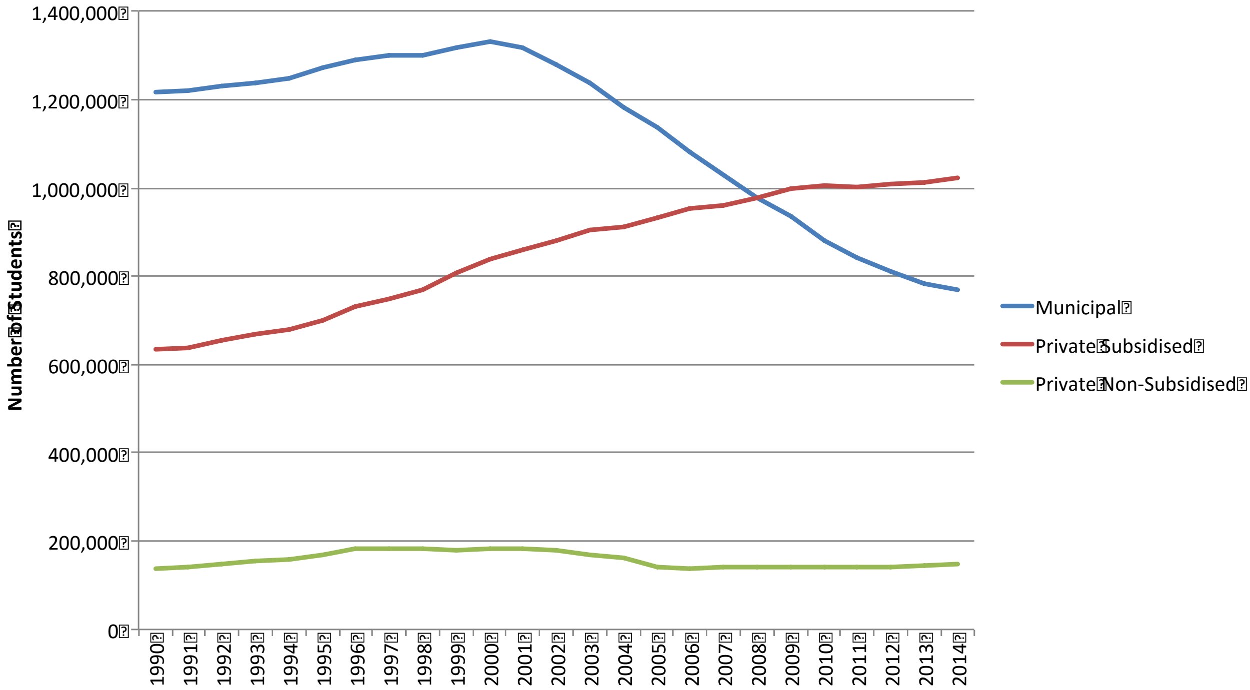
- Municipal (public) schools: this sector is government funded via the attendance-based, per-pupil public subsidy (voucher) and administered by municipal governments.
- Private subsidised schools: these schools are administered by private non-profit or for-profit organisations and financed through both, the voucher system and parental contributions.

- Private non-subsidised schools: these schools administered by private non-profit or for-profit organisations that do not receive public subsidies, and operate entirely through parental contributions.

In 2014, the proportion of student enrolment in each of these categories of schools was as follows: from a total of 1,939,926 students in primary school, 39.6% attended municipal schools, 52.7% went to private subsidised schools and the remaining 7.7% studied in private non-subsidised schools (MINEDUC, 2015a). Figure 2 shows the trend of growth in primary education student enrolment since 1990. While the enrolment in municipal schools has decreased considerably in the last decade, private subsidised schools have experienced an increase in their enrolment of more than 60% since 1990. Also, as Figure 3 indicates, when compared to other systems, the percentage of students attending private schools in Chile is among the highest internationally, with figures similar to those in the Netherlands and Ireland.

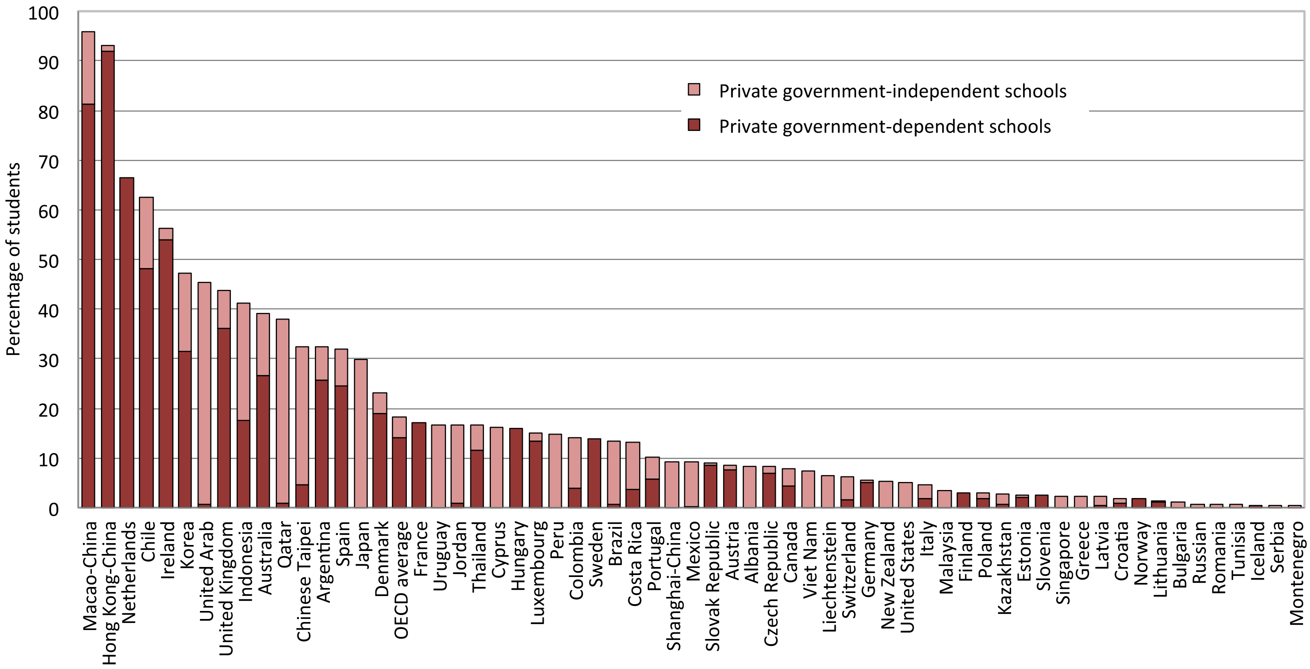
The Chilean system is strongly segregated by social class, with municipal (public) schools serving the poorest families, private-subsidised schools serving the middle class and the elite private non-subsidised schools serving the richest (Mella, 2003). The evidence shows that there are important differences in terms of academic achievement among school sectors, in that private schools, operating in more socially advantaged contexts, systematically attain higher achievement levels than municipal schools, serving more disadvantaged communities. For example, in PISA 2012 the difference in mathematics achievement between public and private schools in mathematics scores was in the order of .65 standard deviations in favour of private school (compared to an OECD total difference of .29 standard deviations) (OECD, 2014a). However, previous research also indicates that the attainment gap in favour of private schools is, to a great extent, explained by differences in the schools' student intake and does not indicate greater educational effectiveness (Bravo, Contreras and Sanhueza, 1999; McEwan, 2001; Mizala and Romaguera, 1998; Mizala and Torche, 2012; Troncoso, Pampaka and Olsen, 2015). Despite this, the school accountability system in Chile still employs and publishes school average scores as indicators of effectiveness (San Martín and Carrasco, 2012).

FIGURE 2: EVOLUTION OF PRIMARY EDUCATION STUDENT ENROLMENT BY SCHOOL TYPE 1990-2014



Source: Department of Studies and Development, Chilean Ministry of Education.

FIGURE 3: PERCENTAGE OF STUDENTS ATTENDING PRIVATE SCHOOLS



Source: OECD (2013), PISA 2012 Results: What Students Know and Can Do, Paris, OECD Publishing.

Most evidence suggests that unrestricted school choice in Chile has exacerbated stratification between public and private schools. Researchers have found that parents search for instructional environments in which their children attend schools with peers of similar social background (Elacqua, Schneider and Buckley, 2006), and that private subsidised schools "skim" off middle and high-income students, while the most disadvantaged students are more likely to attend public schools (Hsieh and Urquiola, 2006). Indeed, students attending any given school are very similar in terms of their socio-economic levels, which puts the system among the least inclusive in the region, together with those of the Dominican Republic, Uruguay, Nicaragua, Costa Rica and Paraguay (UNESCO, 2015b). While school stratification can be explained in part by social and economic inequalities, and by the high levels of residential segregation in the country, schools, through their selection processes, generate further segregation that goes beyond the existent socio-economic inequalities (Elacqua, 2012; Flores, 2008). Thus, as different authors have pointed out (García-Huidobro, 2009; Mizala, Romaguera and Ostoic, 2004), the Chilean schooling system is characterised by social class stratification and segregation.

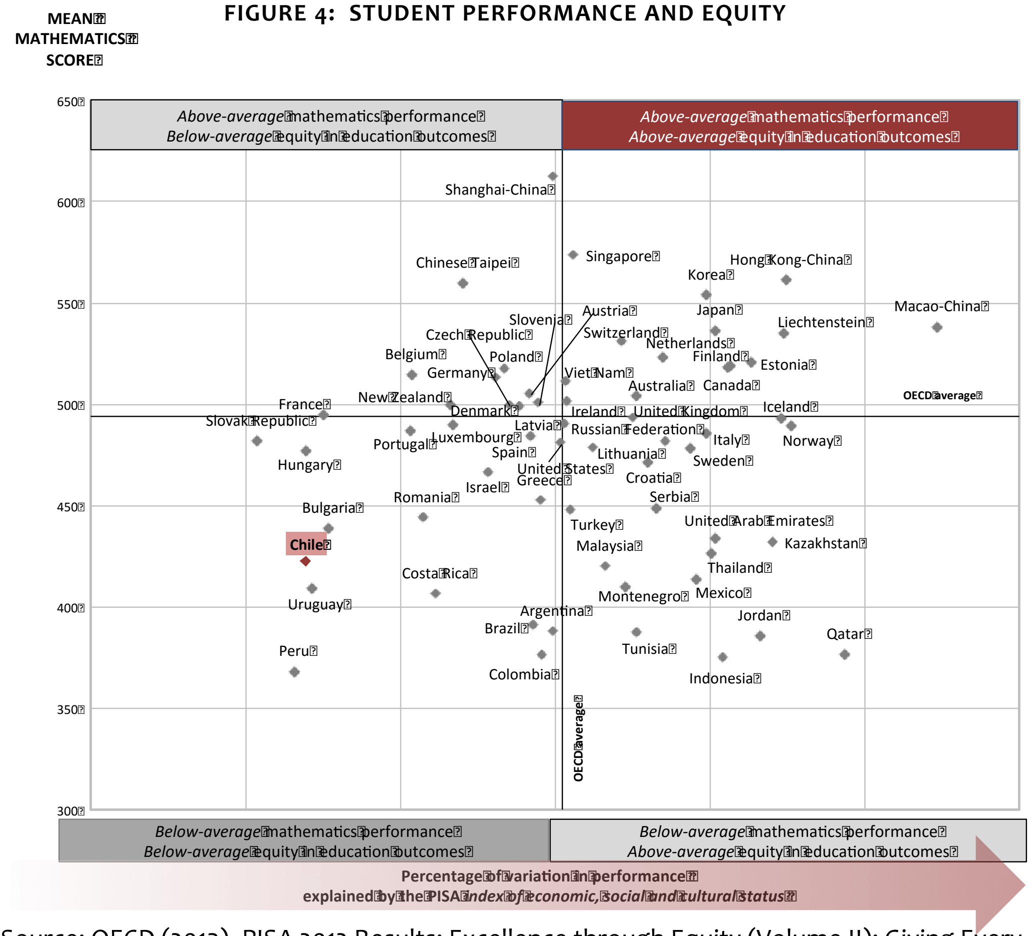
1.2.1.2 Student Achievement in Chile

Chile is among the highest-performing systems in the Latin-American context (UNESCO, 2008b; UNESCO, 2015c). Similarly to Costa Rica and Mexico, its performance in TERCE—the third wave of the UNESCO comparative study—is consistently above the regional average in all of the subjects (reading, writing, mathematics and natural sciences) and grades (Grades 3 and 6) assessed.

However, when compared to post-industrialised economies, the country performs consistently below the OECD average in PISA mathematics, reading and science tests (OECD, 2015b), and below the mathematics and science average scores among the broader sample of countries participating in TIMSS (Mullis et al., 2012). Moreover, the strength of the relationship between student performance and socio-economic status (the proportion of the variation in performance described by the variation in socio-economic status) is above the OECD average: an indication of inequities in education outcomes. Indeed, more than 23.1% of the difference in student performance can be attributed to their socio-economic status, while on average in OECD countries socio-economic status accounts for 14.8% (OECD, 2013a).

The relative position of Chile, in terms of both mathematics performance and the impact of socio-economic status on it, is depicted in Figure 4. When looking at Chile's results in comparison to those of the other 63 countries that participated in the Programme for International Student Assessment (PISA) 2012, it is possible to see that Chilean 15-year-old students not only scored significantly below the OECD average in mathematics, but the impact of their socio-economic status on their performance was one of the largest among OECD countries, suggesting that there is plenty of room for improvement in terms of both overall performance and equity. Thus, while mathematics and reading performance show improvement across SERCE-TERCE, PISA, TIMSS cycles (Mullis et al., 2012; OECD, 2015b; UNESCO, 2015c), achievement levels and equity remain important issues, particularly for students from socio-economically disadvantaged backgrounds.

FIGURE 4: STUDENT PERFORMANCE AND EQUITY



Source: OECD (2013), PISA 2012 Results: Excellence through Equity (Volume II): Giving Every Student the Chance to Succeed, Paris, OECD Publishing.

In Chile, there are also large differences in performance based on gender. The country recorded the largest difference in mathematics performance between boys and girls (favouring boys) in PISA 2012. Furthermore, as evidence from TIMSS shows, the gender difference in mathematics achievement is larger at Grade 8 than at Grade 4 (Mullis et al., 2012). Moreover, it has also been found that the male advantage in Chile varies greatly across schools (Ma, 2008). Conversely, girls generally tend to outperform boys in reading in both, Grades 3 and 6 (UNESCO, 2015b).

1.2.1.3 The Provision of Initial Teacher Training

Teacher education in Chile takes place at the tertiary level, in two types of institutions: universities and professional institutes. The tertiary education level has become a decentralised and unregulated system, where private provision has expanded substantially over the last four decades (Brunner and Uribe, 2007; Matear, 2007). Consequently, the Ministry of Education has little control over the curriculum and arrangements of the programmes offered by teacher training institutions. A growing number of student teachers are being trained in a wide range of public and private institutions, offering different instructional arrangements such as distance learning and evening programmes, and varying significantly in terms of their duration (five years on average) and content focus (Avalos and Matus, 2010). Furthermore, Chile and England are the only countries in the OECD in which a practicum is not a required component of pre-service training for primary teachers (OECD, 2013b). Some training institutions offer supplementary subject-matter specialisation, requiring student teachers to take additional courses and major in a specific academic subject.

The number of students enrolled in teacher training programmes in Chile has more than quadrupled between 1970 and 2010 (Cox, Meckes and Bascopé, 2010). According to the National Council of Education (CNE), there were 54 institutions (universities and professional institutes) in 2015 running 136 different primary teacher education programmes and training 10,625 primary student teachers across the country (Consejo Nacional de Educación, 2015). This accelerated growth in enrolment bears no relation to the demographic trend of decrease in the school-age population.

In Chile, there are no specified entry standards for teacher education other than those in force for entry into university or non-university tertiary education. In practice, this means that students are usually selected into training institutions in accordance with their secondary school grade average, and their scores on the voluntary national university entrance examination (called PSU, which stands for University Selection Test, in Spanish). However, requirements for entering these programmes vary substantially across institutions. While public universities and the most prestigious private universities are more selective, requiring high PSU scores, the rest of private institutions are less selective, and usually only ask students to meet their financial requirements. PSU scores for future teachers are generally lower than those required for professions such as engineering, computing, mathematics, accountancy and economics. The sole requirement for appointment to a teaching position is to have a teaching qualification from a university or a professional institute. There is no other form of certification, nor is there a registration requirement.

One of the main sources of diversity among teachers in Chile is their initial training. Indeed, even over a decade ago, the Organisation for Economic Co-operation and Development (OECD) report on the Chilean education system (OECD, 2004b) emphasised the need to attend to the heterogeneous quality of initial teacher training. Chilean primary teachers at the end of their teacher education showed variable and overall poor results in the mathematics and pedagogy content knowledge TEDS-M tests, where the country ranked 15th out of the 16 participant countries (Tatto et al., 2012). The national diagnostic assessment for recently graduated teachers "INICIA" has also revealed large variation in subject and pedagogical knowledge across training institutions (MINEDUC, 2015b). Furthermore, previous research has shown that initial teacher training programmes in the country differ significantly in terms of their effectiveness in promoting pre-service teachers' pedagogical and subject knowledge, after controlling for student intake (Manzi et al., 2012). However, the relationship between teacher effects on their students' achievement and the different features of their training, such as duration and subject specialisation (major), has not been investigated.

Recent policies, aimed at monitoring and improving the quality of initial teacher education, have only resulted in indirect mechanisms of quality assurance. Assessments of teachers' knowledge and skills are in place at different levels of the teaching career, generally associated with financial incentives. The government has also introduced a system of programme accreditation. While the accreditation of initial teacher training programmes is compulsory, the consequences of not being accredited are not severe (Donoso, 2008; Ingvarson et al., 2013).

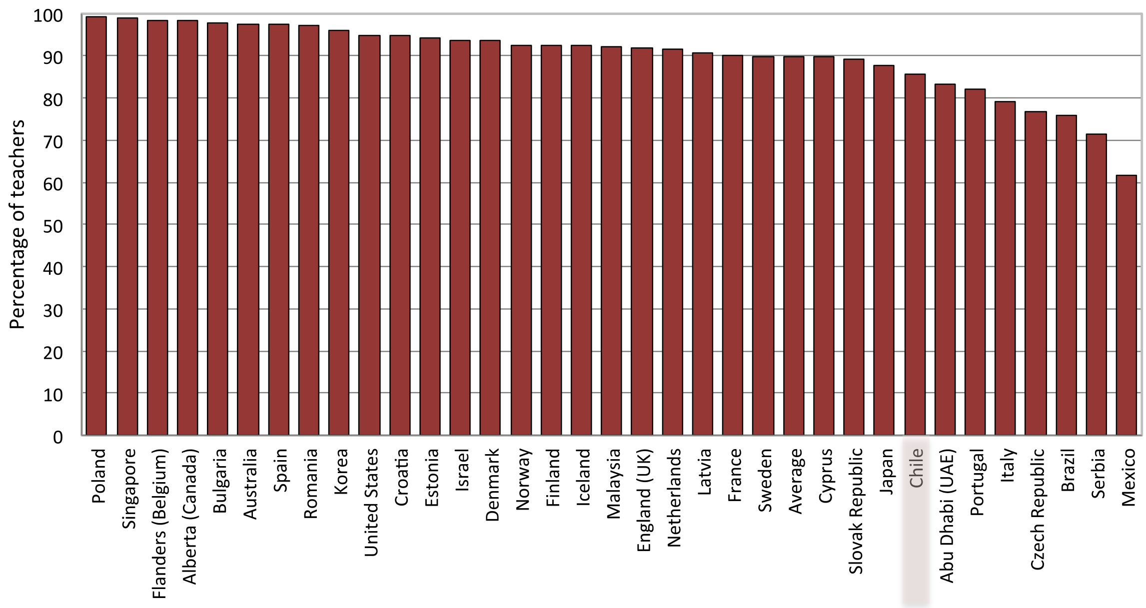
Ingvarson et al. (2013) analyse the quality assurance mechanisms in teacher education in the countries that participated in TEDS-M. The authors find that stronger quality assurance systems are positively associated with teachers' higher mathematics disciplinary and pedagogical knowledge. They also classify the robustness of quality assurance procedures in Chile as weak in comparison to the other participant countries, in all of the different aspects analysed, namely: promotion of teaching as an attractive career, recruitment and selection, accreditation of teacher education programmes and entry to the teaching profession.

1.2.1.4 The Composition of the Teacher Labour Force

This section summarises the main characteristics of the teaching labour force in Chile. According to the most up-to-date data released by the Ministry of Education, in 2014, there were 215,653 teachers in the country. Out of this total, 43.7% worked in municipal schools, 46.1% in private subsidised schools, 9.2% in private non-subsidised schools and 1.0% in other types of schools. Also, 88.4% served in urban schools. In terms of educational levels, 56.4% worked in primary schools, 26.7% in secondary schools and 9.2% in pre-school. Of the rest, 5.6% worked in special education and 2.1% in adult education. Out of the total, 83.1% were classroom teachers, 6.1% were principals or worked in the leadership teams of schools, 3.7% were engaged in technical or other pedagogical roles and the remaining 7.1% performed other roles (MINEDUC, 2015a).

Teaching is a highly feminised profession in Chile, with 74.0% of female teachers in the total system and a higher percentage of 77.6% in the primary level. Despite these high proportions of female teachers, men filled 41.2% of the positions in the leadership of schools. With regards to their qualifications, 94.5% of teachers hold a professional title in education, 3.7% have a professional title in a different area and only 1.8% have not been awarded any type of tertiary education qualification. This makes Chile the country with the highest percentage of certified teachers in Latin America (UNESCO, 2015b). However, as shown in Figure 5, a smaller proportion of teachers than in most OECD countries have completed a teacher education programme (OECD, 2014b).

FIGURE 5: PERCENTAGE OF TEACHERS THAT HAVE COMPLETED A TEACHER EDUCATION PROGRAMME



Source: OECD (2013), PISA 2012 Results: What Students Know and Can Do, Paris, OECD Publishing.

When compared to the other 33 countries that participated in the TALIS study, teachers in Chile had spent fewer years teaching (15.1 years), and were on average younger (41.3 years) than in most other surveyed countries (OECD, 2014b). Also, teachers in Chile report spending 73% of their lesson time on actual teaching and learning. This means that 26% of their time is spent on administrative tasks and keeping order in the classroom (11% and 15%, respectively), which is higher than in other TALIS countries. Finally, teaching conditions in primary and secondary institutions in Chile include teaching time and class size above the OECD average, and below-average salaries, although these indicators vary across urban versus rural locations and school sectors (Santiago et al., 2013).

In the previous sections, the relevance of the study of educational effectiveness and inequality in Chile was highlighted, and the context of the study was introduced. In what follows, the organisation and contents of the remaining chapters of the thesis are presented.

1.3 STRUCTURE OF THE THESIS

The present thesis consists of seven chapters. Following this introductory chapter, which presented the rationale and context of the research project, the remainder of the thesis content is organised as follows:

Chapter 2: A Review of the Literature

In this chapter, the relevant bodies of literature that inform the study are examined. This section addresses the development of the field of educational effectiveness research, and critically evaluates previous empirical research on student achievement trajectories and schools' and teachers' contributions to student achievement. Here, the most common empirical approaches to estimate school and teacher effects are presented, and relevant model specification issues are discussed. Finally, a conceptual framework to guide the study is proposed.

Chapter 3: Research Aims and Methodology

This chapter presents the general methodological approach of the study, covering the following aspects:

- **Aims of the Study:** The three guiding research aims are listed.
- **Research Design:** The general structure of the study, comprising three research phases, is described in this section.
- **Sources of Secondary Data:** The different administrative and student assessment data sets merged to model student achievement trajectories and educational effects are introduced.
- **Data:** The sample resulting from the matching strategy is presented together with the analysis and treatment of missing data. In addition, issues regarding data quality are addressed.
- **Method:** The range of multilevel models implemented and combined in the present study is introduced, and estimation details are discussed.

- **Data Management:** The practical approach to managing and storing the data is presented along with the software packages used for data management and analysis.
- **Ethical Considerations:** A statement related to ethical issues has been included, although this section is kept short as the present research analyses previously anonymised data provided by the MIDE-UC Assessment Centre of the Catholic University of Chile and the Ministry of Education of Chile.

Chapter 4: Student Achievement Trajectories in Primary School

In this section, students' growth in language and mathematics achievement in primary schools and the student-level characteristics that affect these trajectories are studied. The chapter is structured as follows:

- **Introduction:** The aims and organisation of the chapter are presented.
- **Research Questions:** The three research questions that guide the analyses in this chapter are listed.
- **Method of Analysis:** The progression of models fitted in this section is presented with emphasis on the specification of the accelerated growth model.
- **Results:** In this section, the convergence of expected trajectories in the accelerated longitudinal design is tested, and the growth patterns found in the observed data are described.
- **Conclusions:** The main findings in relation to the relevant features of students' achievement trajectories, which are carried forward to the statistical models in the following chapters, are highlighted.

Chapter 5: School Effects on Student Achievement Trajectories

This chapter addresses the main properties of school effects in Chile, which include the magnitude of these effects on student achievement status and growth, school compositional effects, the consistency of school effects across academic subjects (i.e., language and mathematics) and the presence of differential school effects across student groups, defined in terms of student gender and socio-economic status. The chapter is structured as follows:

- **Introduction:** The aims of the chapter are presented.
- **Research Questions:** The four research questions that guide the analyses in this chapter are listed.
- **Method of Analysis:** The three-level models fitted in this section and their connection to the research questions are described.
- **Results:** The different properties of school effects addressed are explored and represented graphically.
- **Conclusions:** The main findings in relation to school effects are summarised and discussed.

Chapter 6: Teacher Effects on Student Achievement Trajectories

This chapter addresses teacher effects in Chile, measured by introducing cross-classification and multiple membership components to the multilevel models fitted in previous chapters. The main aspects to be investigated are: how large are teacher effects on students' achievement status and growth, how consistent are these effects across different academic subjects (i.e., language and mathematics), which teacher input variables explain the variation in teacher effects and whether teacher effects accumulate over time. The chapter is structured as follows:

- **Introduction:** The rationale and research aim of the chapter is presented.
- **Research Questions:** The four research questions that guide the analyses in this chapter are listed.
- **Method of Analysis:** The cross-classified and multiple membership models fitted in this section and their connection to the research questions are presented.
- **Results: T**he different properties of teacher effects addressed are explored and represented graphically.
- **Conclusions:** The main findings in relation to teacher effects are summarised and discussed.

Chapter 7: Discussion and Conclusions

In this final chapter, the findings of the different phases of the research project are discussed, brought together and set in the wider context of the literature, covering the following contents:

- **Main Results and Contributions to the Field of Educational Effectiveness Research:** A discussion of the results presented in each phase will be reported, comparing them with those found in the literature. The contributions of the thesis to the EER field are also discussed.
- **Strengths and limitations of the study:** This section lists the strong aspects of the research projects and its limitations, imposed mainly by the availability and quality of secondary data.
- **Recommendations for further research:** Suggestions for future research are presented in terms of issues that have not been answered in this study and still need attention.
- **Implications for Chilean Education Policy:** The implications of the findings that are particularly relevant to policymakers in Chile are discussed.

Finally, a bibliography and seven appendices follow this section.

CHAPTER 2: A REVIEW OF THE LITERATURE

This chapter summarises some of the main developments in the field of educational effectiveness, critically evaluates previous empirical research on the topics addressed in this study and points out relevant knowledge gaps. Special attention is given to state-of-the-art research on student achievement trajectories and the estimation and properties of school and teacher effects. A conceptual model that synthesizes the issues addressed in this thesis is presented in the end of the chapter.

2.1 AN INTRODUCTION TO EDUCATIONAL EFFECTIVENESS RESEARCH (EER)

This project is situated in a broad tradition of studies investigating school and teacher effects on student outcomes: the field of Educational Effectiveness Research (EER). The field arose as a response to early studies that emphasised the little variance on student achievement that educational factors explained when compared to student background characteristics (i.e., Coleman et al., 1966; Jencks et al., 1973). During a first stage, EER was interested in providing evidence of significant and systematic differences between schools in their abilities to raise student achievement. Thus, using similar quantitative methods, studies conducted during the 1970s by Brookover et al. (1979) in the United States, and Rutter et al. (1979) in England, were concerned with demonstrating that schools could make a difference in students' attainment. These studies, identified as the precursors of the field, were followed by others in different national contexts (Creemers, Kyriakides and Sammons, 2010; Rivkin, Hanushek and Kain, 2005).

The focus of a second generation of studies was to investigate how much schools differ and, furthermore, to what extent they differ when their student intake is similar in terms of innate abilities and socio-economic background (Creemers and Kyriakides, 2008). However, contemporary EER does not solely assesses differences between schools in terms of effectiveness, but it is also concerned with theorising why some schools and teachers are more effective than others, when the differences in performance cannot be explained by the composition of schools or classrooms in terms of student characteristics. As Creemers and Kyriakides (2008, p. 12) state, *"the main research question of EER is what factors in teaching, curriculum, and learning environment at different levels such as the classroom, the school, and the above-school levels can directly or indirectly explain the differences in the outcomes of students, taking into account background characteristics, such as ability, SES, and prior attainment"*. Moreover, EER seeks to link these findings to school improvement (Creemers, Kyriakides and Sammons, 2010).

Reviewers identify three main disciplinary perspectives that have shaped the field (Teddlie and Reynolds, 2000). In a first stage, EER borrowed from economics and followed a common framework of statistical inquiry based on production functions to discover determinants of school and teacher effectiveness, by investigating the effects of varying primary inputs (such as school expenditure, student-teacher ratio and teacher qualifications) on student achievement (output) (e.g., Hanushek, 1997; Monk, 1992).

More recently, EER has gradually shifted its interest towards the impact of process variables (the educational "black box"), as research based on the input-output approach has not provided consistent relationships between variation in school resources and student performance (Teddlie and Reynolds, 2000). However, in developing countries, school input variables seem to have a greater explanatory power than in post-industrialised economies. Indeed, researchers have suggested that the impact of resources or inputs on educational outcomes play a major role in regions such as Latin America and the Caribbean (Farrell and Oliveira, 1993; Fuller and Clarke, 1994; Hanushek, 1994; Hanushek, 1995; Heyneman and Loxley, 1982; Scheerens, 2000; Velez, Schiefelbein and Valenzuela, 1993; Willms and Somer, 2001). The authors attribute the greater impact of school inputs in developing countries, compared to industrialised countries, to the greater variance in both school inputs and outputs.

A second influential tradition in EER is sociology. One of the key contributions of the sociological perspective on EER has been to stress the importance of assessing the impact of input factors concerned with the educational background of students, such as their socio-economic status and cultural capital, on student achievement. This area of research has also pushed towards the analysis of to what extent education compensates for prior achievement differences (by adapting to the needs of different groups of students) or, instead, reproduces or increments achievement gaps among students (Campbell et al., 2004; Kyriakides, 2004). This consideration has led to an expansion of the criteria for measuring effectiveness, which not only refers to promoting the general achievement progress of a group of students (excellence dimension), but also to reducing the variance in student outcomes (equity dimension). As it will be explained later in this chapter, this line of inquiry has provided evidence against generic educational effects and explored the existence of differential educational effectiveness (Creemers and Kyriakides, 2008). The sociological perspective has also promoted the use of process variables borrowed from organisational theories, such as school climate, culture and structure, and highlighted the relevance of contextual and compositional factors (Opdenakker and Van Damme, 2006).

Finally, as mentioned above, psychology has also influenced the field of EER to a great extent, by stressing the importance of studying the learning processes that take place inside the classroom, paying more attention to the learning and instructional level than previous research in the field.

Efforts to systematise the findings of the vast and diverse research literature in the field have led to the development of theoretical models of educational effectiveness. Most notably, the Dynamic Model of Educational Effectiveness, introduced by Creemers and Kyriakides (2008) has emphasised on the changing nature of educational systems, and the complexity of relationships between factors operating at different levels. This conceptual model places the study of change at its heart, and highlights the multilevel nature of factors affecting student achievement.

Important theoretical and methodological advances have allowed the field to grow and diversify. Indeed, different reviewers agree in pointing out that over the past three decades, EER has undergone important advances in the methodological arena (Creemers, Kyriakides and Sammons, 2010; Goldstein, 2011). In light of these developments, two main lessons can be learnt: acknowledging the multilevel structure of educational processes, and collecting longitudinal data for the study of learning

(Creemers and Kyriakides, 2006). These key aspects of educational effectiveness research are discussed below.

2.1.1 *The Multilevel Analysis of Educational Effects*

A methodological imperative in the measurement of educational effects is to pay attention to the multilevel organisational structure in which education occurs (Creemers and Kyriakides, 2008). When working with data from a sample composed of two or more levels of aggregation, it is highly likely that observations will not be independent from each other. For example, students within classes or schools are more likely to share common characteristics than those from different classes or schools. Traditional statistical analyses generally assume the independence of observations and, when this assumption is not met, the standard error is underestimated, leading to statistically significant results that are spurious (Walsh, 1947). Multilevel analysis deals with this issue by incorporating the dependency of observations into designs that can include a range of variables measured at different levels (Goldstein, 2011; Hox, 2010; Raudenbush and Bryk, 2002; Snijders and Bosker, 2012). In other words, multilevel models assume that students are not randomly assigned to teachers, and that teachers are not randomly distributed across schools. Furthermore, multilevel analysis allows to simultaneously study the influence of factors at different levels, which avoids ecological and atomistic fallacies by permitting associations between two variables at the group level (or ecological level) to differ from associations between analogous variables measured at the individual level.

From a substantive point of view, multilevel analysis provides a means of partitioning the variance of the outcome variables into different levels. This, in turn, allows us to estimate educational effects, such as school and teacher effects (Creemers and Kyriakides, 2008). As several reviewers have concluded, the detailed analysis of variation in students' educational outcomes, and the way various sources of influence help to shape students' learning and developmental outcomes over time, is well supported by multilevel modelling, which has become a powerful tool for EER (Luyten and Sammons, 2010; Rumberger and Palardy, 2004). Indeed, as it will be discussed in following sections of this chapter, multilevel analysis provides the basis for value-added models developed to measure school and teacher effects.

In recent years, the use of multilevel analysis in the field has become more sophisticated, improving the modelling of measurement error and cross-level interactions. The use of advanced multilevel models has also supported the development of more complex value-added models by specifying additional levels of variation, such as neighbourhoods, classrooms, primary and secondary schools, and local education authorities, which have been found to be significant sources of variation in student achievement (De Fraine et al., 2003; Goldstein, Simon and Brendon, 2007; Leckie, 2009; Martínez, 2012; Rasbash et al., 2010; Timmermans, Snijders and Bosker, 2013). Importantly, as it will be discussed in later sections, cross-classified and multiple membership models have been applied to not perfectly nested data structures, such as those encountered in longitudinal educational research following students across classroom settings, to assess teacher and school effects (Beretvas, 2011; Raudenbush and Bryk, 2002).

2.1.2 *A Longitudinal Approach for the Study of Learning*

Research looking at the complex phenomenon of student learning is often limited in that it stems largely from cross-sectional designs or two time point longitudinal designs. Cross-sectional designs have both statistical and conceptual limitations. Statistically, a cross-sectional design confounds age and cohort differences by not being able to separate different sources of variations. Conceptually, a cross-sectional design does not reflect the fact that "*the very notion of learning implies growth and change*" (Willett, 1988, p. 346), because it assumes that student achievement—the outcome of learning—can be represented as a score measured at a single time point, instead of scores measured at multiple time points to show a growth pattern. Two time point longitudinal designs are also unsatisfactory, as they provide a very limited source of intra-individual variability to study change in student achievement (Bryk and Raudenbush, 1987).

Educational research has found that a longitudinal approach has been useful when looking at the dynamics of learning. Longitudinal data are also needed to measure the effect of schools and teachers on student achievement growth, as they themselves are changing entities.

An important advantage of longitudinal data is that it can enhance the validity of causal inferences in non-experimental research, by providing a basis for assessing the direction of causation between variables, and by making possible some control over selection effects.

The criteria for satisfactory inference in studies of educational effectiveness have often included a longitudinal design with repeated measures on multiple cohorts (Goldstein, 1997; Gray, 1995; Sammons, Thomas and Mortimore, 1997). As longitudinal data become more widely available, researchers have tended to shift their focus from achievement to learning, with the latter involving changes over time usually studied using growth curve models (Raudenbush and Bryk, 2002; Scheerens and Bosker, 1997; Singer and Willett, 2003). However, growth curve analysis, allowing the comparison of progress across more than two time points, is still infrequent in educational effectiveness research models[8].

Longitudinal data are multilevel in nature, as measurements are grouped within subjects. Indeed, a common application of multilevel models is the analysis of growth curve data in repeated measures designs. In this context, measurement occasions are set as the lowest level of analyses, grouped in turn into students, teachers, schools, etc. Raudenbush and Bryk (2002) and Hox (2010) identify the key advantages of using multilevel models to analyse repeated measures data, namely, that by modelling varying regression coefficients at the measurement occasion level, different growth curves for each individual subject can be estimated resembling their individual development. In addition, the covariances between the repeated measures can be modelled, by specifying a specific structure for the variances and covariances, allowing us to investigate the relationship between level (intercept) and change (slope) over time.

Furthermore, hierarchical linear models for repeated-measures data admit adding higher levels to investigate the effect of groups (such as school) on differences in individual development (e.g., Guldemond and Bosker, 2009; Van de gaer et al., 2009). Finally, it is also straightforward to include time-varying or time-constant explanatory variables in the model, which allows us to model both the average group development and the development of different individuals over time. As growth curve models assume that there is a latent growth trajectory that underlies the observed outcomes, they provide a

[8] The researcher conducted a systematic search and 56 journal articles were found that made use of growth curve models to estimate school and/or teacher effects.

more reliable measure of students' progress, and increased power for detecting statistical effects (De Fraine et al., 2005). Finally, many researchers applying multilevel analysis to the study of student achievement growth prefer polynomial growth models, rather than simple linear growth models, in order to capture nonlinear trends.

In the following sections, previous research on student achievement growth trajectories and school and teacher effects is discussed.

2.1.3 Student Achievement Trajectories

As mentioned above, a longitudinal approach is being promoted and regarded as a key criterion in EER, as it is suitable for the study of pedagogical processes and their association with change in student attainment (Creemers, Kyriakides and Sammons, 2010; Goldstein, 1997; Hill and Rowe, 1998; Raudenbush, 1989; Teddlie and Reynolds, 2000).

When researchers have studied student achievement trajectories over time, these have been found to follow a curvilinear shape rather than a linear trend in a wide range of national contexts (e.g., Brooke and Bonamino, 2011; Caro et al., 2009; Guldemond and Bosker, 2009; Muthén and Khoo, 1998; Verhaeghe, Van Damme and Knipprath, 2011). Furthermore, it has been found that student academic progress is usually more accelerated in primary school, and becomes less steep towards the transition to secondary education (Hill and Russell, 1999; Kiplinger, 2004; Kolaweski-Jones and Duncan, 1999; Zvoch and Stevens, 2003). While there is a wide range of student-level factors affecting educational achievement, the present study, and therefore this review of the literature, focuses on two of the most commonly researched variables: gender and socio-economic status (SES).

2.1.3.1 The Gender Gap

Student gender has been considered an important predictor of student achievement at both primary and secondary levels. International cross-sectional research has often found gender differences in academic achievement across educational contexts. Indeed, studies using large-scale assessment data have revealed a *"gender division of learning"* (Ma, 2008) as gender differences in favour of females are consistently found in language performance in almost all countries, with systems usually demonstrating from moderate to large gender gaps. Conversely, consistent gender differences favouring boys in mathematics performance have been found in most countries, with gender gaps being generally small (Ma, 2008; Marks, 2008; Mullis et al., 2000; OECD, 2013a; Treviño et al., 2010).

While overall gender differences in performance are well documented in the literature, it is still unclear whether gender gaps remain stable, widen or narrow throughout students' academic trajectories over their school years. Indeed, little attention has been paid to gender differences affecting achievement growth, and the scarce existent research, found mainly in the field of developmental psychology, is inconclusive on whether achievement growth rates differ between genders.

Moreover, most studies have addressed growth in achievement at the latest stages of primary school and at the secondary school level (probably because it is at these stages when gender differences in achievement levels appear or become more pronounced in post-industrialised countries). For example, using a latent growth curve approach, Muthén and Khoo (1998) study mathematics achievement developing over grades 7 to 10 in two cohorts of students using data from the Longitudinal Study of American Youth (LSAY). The authors found no significant gender difference in achievement growth trajectories. Subsequent studies have also failed to find significant gender gaps in mathematics achievement growth rates in later primary and early secondary school (e.g., Ai, 2002; Mok et al., 2015; Wilkins and Ma, 2002; Willms and Jacobsen, 1990). Similarly, but looking at language achievement, Cole (1997) concludes that gender differences remain relatively stable between Grade 8 and Grade 12.

These results conflict with those of De Fraine, van Damme and Onghena (2007). The authors found gender differences in language achievement trajectories from Grade 7 to Grade 12 in Flanders, as girls had an increase in language achievement over time, whereas boys showed a decrease in middle years, followed by an increase from Grade 9 onwards. Kiplinger (2004) also suggests that female secondary school students progressed at a significantly faster rate in reading achievement than male students. Looking at mathematics achievement, Leahey and Guao (2001) found that, despite relatively equal achievement levels in primary school, boys' growth accelerates faster than girls', leading to small gender differences in secondary school. Likewise, Vasquez-Salgado and Chavira (2014), suggest that Latino boys and girls in the United States follow different growth trajectories in mathematics, but only in secondary school.

Previous research has also provided other interesting findings regarding gender differences in achievement growth. For example, Kolaweski-Jones and Ducan (1999) find that boys and girls have differing amounts of variability in their trajectories, with boys exhibiting greater variability in the slopes of their achievement growth. Also, gender gaps in achievement growth might vary significantly across schools and among student groups. For example, Ai (2002) finds that gender differences in growth in mathematics vary across students with different initial achievement status in mathematics: while achievement growth for students with high initial status showed no gender differences, for those who started low, girls started higher than boys, and their average growth rate was slightly lower than boys.

Several reasons could explain differences in growth trajectories of male and female students, including differential rates of cognitive development, different growth in attributing academic success to strategy use (Mok, Kennedy and Moore, 2011), and differing enjoyment and attitudes towards academic subjects, which in turn leads to differential effort expenditure (Pinxten et al., 2014). However, research has also suggested that families and schools are chief contributors to gender gaps in academic achievement, particularly through differences in parents' and teachers' encouragement for boys and girls (Ai, 2002).

In conclusion, the existing literature on gender differences examines mainly students' achievement status rather than rate of growth. The evidence suggests that, while gender differences in students' achievement status in language and mathematics may be a global phenomenon, differences in students' rate of growth in mathematics seem to be a local phenomenon (Ma, 1999). A longitudinal perspective, looking at achievement trajectories as developmental processes, can shed light on how gender gaps progress. Also, most research in this topic has been conducted in the United States and other developed countries. Therefore, more longitudinal evidence is, therefore, needed; particularly studies looking at gender differences in achievement growth at the primary school level, which, in many developing countries, is the stage at which gender differences in achievement levels originate or are already established.

2.1.3.2 *The SES Gap*

Education inequalities are often measured through the educational attainment gap between pupils from affluent and socio-economically disadvantaged families (Caro and Cortés, 2012). The achievement gap between high- and low-socioeconomic status (SES) students, well documented in the literature (Sirin, 2005; White, 1982), manifests in school but emerges earlier in life.

Previous research has also studied how the achievement gap related to family SES changes as students progress through school. This is of critical importance to policy because it can offer insights into how and when inequalities reproduce, and into the particular policies that can contribute to reducing disparities among SES groups during the school period. Here, most studies anticipate a widening achievement gap among SES groups as students advance in school: a pattern known as the Matthew effect (or accumulated advantage) where "the rich get richer and the poor get poorer". The literature offers several theoretical reasons for this phenomenon, which can be broadly divided into: those concerned with the consequences of structural locations of students in schools, derived from tracking and streaming practices (Condron, 2007; Gamoran, 1992; Haller and Davis, 1981; Kerckhoff, 1993; Oakes, 1985), those related to the effect of family resources and, more generally, the out-of-school environment during the summer break (Alexander, Entwisle and Olson, 2001; 2007; Downey, von Hippel and Broh, 2004; McCoach et al., 2006) and those associated with the mechanisms whereby students regulate individual effort in school (Breen and Goldthorpe, 1997; Goldthorpe, 1996; Guo, 1998).

While the literature suggests that the situation of low-SES children tends to worsen as they get older, empirical evidence is limited in that it stems largely from cross-sectional designs or two time point longitudinal designs (e.g., Gamoran, 1992; Guo, 1998; Willms, 2002). As discussed above, the former confound age and cohort effects, and the latter provide a very limited source of intra-individual variability to study change in the gap (Bryk and Raudenbush, 1987). Some research, however, relies on sounder designs and statistical methods (e.g., Alexander, Entwisle and Olson, 2001; Caro, McDonald and Willms, 2009; Caro and Lehmann, 2009; Downey, von Hippel and Broh, 2004; Wilkins and Ma, 2002). These studies use a multilevel approach applied to three or more time point longitudinal designs, which enables the separation of student and school variation from variation due to test-level measurement error, while drawing on a substantial source of intra-individual variation to examine change.

Most previous research indicates a widening SES gap over the course of primary school, which unfolds mainly during summer breaks. However, results are not conclusive with regards to the specific timing at which differences in achievement growth develop. For example, Alexander et al. (2001) finds that achievement growth rates are greater for high-SES students than for low- SES students during the first 2 years of primary school and level off thereafter. Caro et al. (2009a), in turn, indicate that the SES gap remains fairly stable during the primary school years except for the latest school grades, when it widens.

Interestingly, both Caro and Lehmann (2009a) and Wilkins and Ma (2002) found a positive relationship between SES and initial status but a negative or non-existent with achievement growth, which would suggest that students of higher SES families exhibit similar rates of growth to those of lower SES students, or even that the SES gap tends to narrow. The authors do not rule out, however, that this is a "learning curve" effect. Namely, higher SES students start near their peak of growth, whereas lower SES students, starting far below the peak, exhibit faster growth rates.

From this review, it is concluded that more longitudinal research that captures a broad range of school years is needed in order to advance our knowledge of the development of the SES achievement gap. In the following sections, the literature addressing the influence of schools and teachers on student achievement is discussed.

2.1.4 School Effects

Several methodological approaches have been applied to estimate the effect of schools on student achievement. EER has often stressed that a school is not responsible for the absolute level of student achievement so much as for the progress made by pupils in its care and, therefore, a measure of the relative gain in achievement made by pupils is required in order to compare schools (Teddlie and Reynolds, 2000).

Value-added models (VAM), first popularised by the work of Sanders and colleagues (Sanders and Rivers, 1996; Sanders, Wright and Horn, 1997), measure the gain (or loss) of being assigned to a given school and/or teacher, and are based on student progress (Raudenbush, 2004). They therefore require at least one lagged measure of the score representing a baseline, and the progress of students in each school and/or classroom are then compared jointly. In other words, the gain (or loss) of being educated in/by a given school/teacher is calculated with respect to an “average” school/teacher (Raudenbush, 2004; Raudenbush and Willms, 1995). Although many types of outcome might be considered (e.g., grade completion or retention, attendance, affective outcomes), most value-added modelling to date has focused on scores from standardised assessments measuring cognitive outcomes (Creemers, Kyriakides and Sammons, 2010).

VAM approaches generally infer effectiveness from a residual estimate of the school or teacher’s current class achievement obtained through statistical adjustment (Kupermintz, 2003). A residual, defined as the observed score minus the “expected” score, is an adjusted estimate after removing the effects of other extraneous sources of variation. In these models, the overall magnitude of school/teacher effects are estimated using the intra-class correlation (ICC): the proportion of variance in achievement scores found at the school/teacher level with respect to the total variance across levels.

It has been frequently noted that the validity of residual estimates as a way of describing school and teacher effectiveness depends to a great extent on the statistical models from which they are derived (Heck, 2009; McCaffrey et al., 2003; Rowan, Correnti and Miller, 2002). Two of the most common empirical approaches to value-added estimates of educational effects on student achievement are covariate adjustment models and gain scores models. The covariate adjustment model uses prior achievement scores as covariates in models for later outcomes (McCaffrey et al., 2004). Such a model has frequently been used in much of the economics production function literature (Hanushek, 1992), and is often applied with only two years of data, providing only one year of school or teacher effect estimates. Gain scores models, in turn, are suitable when all scores are on the same scale, and scores from adjacent grades can be subtracted to obtain "gains" that become the dependent variable in the analysis (Rowan, Correnti and Miller, 2002)[9].

There is considerable debate over the use of these models. As argued by Rowan et al. (2002) both approaches are likely to underestimate school and teacher effects on student achievement, due to the inclusion of test-level measurement error in the estimation of the student-level variance, which, in turn, inflates the total variance against which school- and teacher-level variance proportions are calculated. Growth models, that directly model the entire vector of student outcome scores jointly, are preferred for estimating value-added scores, as they yield estimates with better properties than covariate-adjustment and gain models (McCaffrey et al., 2003; Sanders, 2006).

Another issue that has been raised in the debate is whether VAM should control for student-level confounding factors such as SES, ethnicity and gender, as well as compositional effects, giving way to contextualised value-added models (CVAM) (Foley and Goldstein, 2012). Some researchers argue that value-added scores are stable regardless of the inclusion of these student covariates, because the inclusion of prior-year test scores accounts adequately for student characteristics and allows students to serve as their own controls (Ballou, Sanders and Wright, 2004). Others, however, claim that failing to adjust for student covariates may be unfair to schools and teachers of disadvantaged students, and it is likely to overestimate educational effects (Amrein-Beardsley, 2008; McCaffrey et al., 2003; Seidel and Shavelson, 2007). CVAMs aim to increase equity in comparisons, as judgments about school and teacher effectiveness are expected to exclude those factors that lie beyond their control (Bingham, Heywood and White, 1991; De Fraine, Van Damme and Onghena, 2002; Salganik, 1994; Timmermans, Doolaard and de Wolf, 2011; van Ewijk and Sleegers, 2010b).

[9] Gain score models are less frequent in research than covariate adjustment models as educational achievement variables are not often standardised across year groups and over time.

2.1.4.1 *The Magnitude of School Effects*

A number of studies have sought to quantify the size of school effects on student outcomes. In a meta-analysis of school effects, Scheerens and Bosker (1997) estimate that, on average, schools account for 19% of the variance in unadjusted student achievement levels, and 8% when adjusting for initial differences between students. A cross-national analysis of mathematics achievement of 15-year-olds using data from the Programme for International Student Assessment (PISA) 2012 shows that the between-school variance ranges from over 50% in countries where formal school tracking begins at an early age to around 10% in Scandinavian countries. The estimate for Chile was 33%, declining to 13% when taking into account students' and schools' socio-economic status (OECD, 2013a).

As explained above, educational effects are more accurately estimated when achievement progress over time is studied (Kyriakides and Creemers, 2008; Raudenbush and Bryk, 2002; Rowan, Correnti and Miller, 2002). The magnitude of school effects in longitudinal studies, where measurement error is controlled to a greater extent, tends to be larger than when estimated using cross-sectional or two time-points designs (leading to covariate adjustment and gain scores models) (Dumay, Coe and Anumendem, 2013; Raudenbush and Bryk, 2002; Teddlie and Reynolds, 2000).

Indeed, the application of three-level growth models has provided strong evidence on substantive variation in student achievement growth that can be attributed to schools in a wide range of contexts (e.g., Guldemond and Bosker, 2009; Rowan, Correnti, & Miller, 2002; Zvoch & Stevens, 2006). For example, Bryk and Raudenbush (1988) demonstrate the estimation of growth trajectories for students and schools and find that in reading and mathematics learning growth rates, 44% and 83% of the variance, respectively, lies between schools.

Also, when comparing the evidence on school effects produced in different national contexts using "raw" or adjusted achievement, the levels of between-school variance in developing countries has been found to be considerably higher than in post-industrialised countries (Riddell, 1997; Scheerens, 2001; Willms & Somer, 2001). This phenomenon seems to reflect greater variability and influence of school resources, such as trained teachers, textbooks and materials in the former. Indeed, Scheerens (1999) explains that this variation across countries may be due to school-level factors showing their effects less strongly when the educational system of a developing country begins to resemble educational systems of industrialised countries.

It follows that the magnitude of school effects on student achievement growth should also be larger than that found in emerging economies. However, most research on school effects using a longitudinal perspective comes from post-industrialised countries, due to the installed capacity to produce large data sets of the quality required for providing valid and reliable estimations. Longitudinal data in developing countries is scarce and is not usually analysed under this approach.

Previous studies also suggest that there is variation in the size of school effects for different academic outcomes, which are larger for subjects such as mathematics and science that are typically learned at school and where exposure is limited in the family and the community, as compared with school effects for language (Mortimore et al., 1988; Scheerens and Bosker, 1997; Thomas et al., 1997b). Finally, school effects have also been found to be larger in the latter stages of schooling (Murillo and Roman, 2011; Sammons et al., 2008; Scheerens and Bosker, 1997).

2.1.4.2 School Compositional Effects

The impact of between-school differences in student intake on student achievement has been a popular research topic in the field of educational effectiveness research. However, there is still disagreement in relation to the nature and magnitude of school compositional effects (Thrupp et al., 2002).

The school SES composition has been shown to play a role in educational attainment in addition to family SES (van Ewijk and Sleegers, 2010a). Several studies have shown that low SES students perform worse in school than high SES students, both because they come from low SES families and also because they tend to attend low SES schools (e.g., Caldas and Bankston, 1999; Caro and Lenkeit, 2012; Rumberger, 2005; Thrupp, 1999; Willms, 1985). This cumulative source of disadvantage is usually referred to as double jeopardy (Willms, 2010). School compositional effects are particularly relevant in education systems with a high degree of SES segregation between schools (Timmermans and Thomas, 2015). That is the case for Chile, where low SES students tend to be clustered in low SES public schools due to system level properties, such as school selectivity, school choice and urban segregation (Bellei et al., 2003; Elacqua, 2012; Valenzuela, Bellei and de los Ríos, 2014).

In addition to the socio-economic makeup of schools, other key school compositional features examined in the literature are aggregated levels of achievement and school heterogeneity in terms of performance. The majority of these studies have established that it is generally beneficial for all students to be part of a school with a high average achievement levels (e.g., Cheung and Rudowicz, 2003; Kang, Park and Lee, 2007; Opdenakker et al., 2002; Peetsma et al., 2006; Willms, 1985) while evidence on the effect of attending cognitively homogeneous schools is mixed and still inconclusive (Hanushek et al., 2003; Rodríguez Menés and Donato, 2015).

However, the measurement of school composition poses significant methodological challenges, as the derived effects can be artifacts of poorly specified data at the individual level (Harker and Tymms, 2004; Televantou et al., 2015). Furthermore, school composition research has often relied on cross-sectional data, which cannot disentangle the effects of school composition prior to any process occurring in schools (school composition effect associated with initial achievement status) from the school composition effects that affect those processes (school composition associated with learning gains) (Verhaeghe, Van Damme and Knipprath, 2011).

Previous longitudinal studies investigating school composition have demonstrated differential effects on achievement on the first measurement occasion and subsequent learning growth. Most of these studies have found significant school composition effects on achievement on the first measurement occasion and no, or considerably smaller, effects on learning growth (e.g., Belfi et al., 2013; Guldemond and Bosker, 2009; Luyten, Schildkamp and Folmer, 2009). More longitudinal research is needed in order to draw valid conclusions regarding school composition effects.

2.1.4.3 *The Consistency of School Effects*

Another property of school effects commonly investigated is their consistency across academic subjects. Meta-analyses indicate that the correlations in "raw" school effects across subjects tend to be around 0.7, 0.8, or even 0.9, but are substantially lower for value-added school effects (Scheerens and Bosker, 1997). All in all, school effects based on value-added models show a moderate degree of consistency across subjects (Luyten, 1994; 1998; Marks, 2015; Sammons, Mortimore and Thomas, 1996; 1997; Scheerens and Bosker, 1997; Thomas et al., 1997b; Willms and Raudenbush, 1989). For example, Luyten (2003) summarises previous studies on the consistency of school effects across language and mathematics, finding a median value of .43. This indicates that effective schools in one curricular area are not nessesarily effective in other areas. However, studies investigating this issue are still scarce, and reviewers have often highlighted the need for further longitudinal research to address the consistency of school effects on achievement growth across subjects, whilst making adequate controls for student intake.

2.1.4.4 *School Differential Effectiveness*

Generic school effects are conceived as the overall impact of the school for an "average" student. However, as previously mentioned, school effects can vary across students as schools may be more effective for one group of students than for another group, which is highly relevant to the concept of equity in education. Education researchers have investigated this issue in studies of what has become known as differential school effectiveness research (Nuttall et al., 1989; Sammons, Nuttall and Cuttance, 1993).

Previous studies have investigated differential school effects in terms of students' prior achievement, gender, ethnicity, and socioeconomic status (Scheerens and Bosker, 1997). With regard to gender, some studies have found that schools are equally effective for boys and girls (Kyriakides, 2004; Sammons, Nuttall and Cuttance, 1993; Thomas, 2001), whereas others found that the gender gap differs from school to school (Strand, 2010). Also, most studies have found none or only modest differential effects with regard to student socioeconomic status (Kyriakides, 2004; Sammons, Nuttall and Cuttance, 1993; Strand, 2010; Thomas, 2001; Thomas et al., 1997a), indicating that more effective schools do tend to improve the attainment of all students, but that they do not remove overall patterns of difference related to students' backgrounds.

2.1.5 *Teacher Effects*

A vast body of EER literature has addressed the question of whether teachers differ in their effectiveness in promoting students' academic achievement. Researchers have used a variety of analytic procedures to estimate the overall magnitude of teacher effects on student achievement, and these alternative procedures have produced markedly different conclusions about this question.

When studying teacher effects from a non-experimental research approach, it is important to consider that teachers are neither distributed randomly among schools nor within schools. The estimated effect of teachers would be biased if the allocation of teachers and students into schools and classes induce a correlation between teacher characteristics and unobserved variables that impact student achievement. This selection effect refers to the idea that if better-qualified teachers tend to teach in more affluent schools due, for example, to the possibility of experiencing better working conditions, then a simple analysis based on unconditional models and cross-sectional data would yield an overestimation of teacher effects on student achievement.

Since research based on cross-sectional data is not likely to overcome this issue (Clotfelter, Ladd and Vigdor, 2006; Hanushek, 1997), more recent literature analyses the impact of teachers in promoting student academic achievement using longitudinal data. This approach follows student achievement in time and identifies the teachers who taught them in each stage, allowing researchers to separately identify the contribution of schools, teachers and students to student achievement over school years (Rivkin, Hanushek and Kain, 2005).

As with school effects, two of the most common empirical approaches to producing value-added estimates of teacher effects on student achievement are covariate adjustment models and gain scores models (McCaffrey et al., 2004). Apart from the shortcomings that also apply to the estimation of school effects, as was discussed above, the structure of these models does not capture the complex structure of relationships between students and teachers, as pupils can be taught by a different teacher each year. In this context, student outcomes do not follow the traditional nested designs of hierarchical models, and alternative model formulations are necessary. Estimating the percentages of variance in achievement growth rates lying among teachers within schools over time based on a growth modelling framework requires the use of cross-classified random effects models (Raudenbush, 1995).

The development of cross-classified multilevel modelling techniques, and its implementation in statistical packages, has allowed the expansion of teacher effectiveness research in recent years (Creemers, Kyriakides and Sammons, 2010). However, these models are still only rarely used in educational research (Beretvas, 2008; Luo and Kwok, 2012), often because of the lack of longitudinal data over several time points. Furthermore, as shown in Table 1, very few statistical models in past studies have accounted for crossed grouping factors in the data when estimating teacher effects.

TABLE 1: SYNTHESIS OF STUDIES ESTIMATING TEACHER EFFECTS ON STUDENT ACHIEVEMENT GROWTH USING CROSS-CLASSIFIED RANDOM EFFECTS MODELS

Study	Country	Number of Cohorts	School Level	Grades	Schools	Teachers	Student	Number of Occasions
Raudenbush and Bryk (2002)	USA	1	Primary School	Grades 1 to 4	-	1,553	3,250	4
Rowan et al. (2002)	USA	2	Primary School	Cohort 1: Grades 1 to 3 Cohort 2: Grades 3 to 6	138 - 166	1,378 - 2,033	5,454 - 6,153	3
Kyriakides and Creemers (2008)	Cyprus	1	Primary School	Grades 1 to 4	28	61	1,681	5
Palardy (2010)	USA	1	Pre-Primary and Primary School	Beginning and end of kindergarten and Grade 1	Not specified	1,553	3,250	4

The first application of the cross-classified model that involved repeated measurements for estimating teacher effects from repeated measurements of student achievement (where students encounter multiple teachers over time) was proposed by Raudenbush (1993). This important contribution to the literature on teacher effects estimation was originally constrained to two levels, and was later extended by Rowan, Correnti, and Miller (2002) to include a third level, schools, wherein both students and teachers are nested.

2.1.5.1 *THE MAGNITUDE OF TEACHER EFFECTS*

Under the covariate adjustment model approach, it has been found that 4% to 16% of the variance in students' adjusted reading achievement, and 8% to 18% of the variance in adjusted mathematics achievement, lies among classrooms (depending on the grade at which the analysis are conducted) (Rowan, Correnti and Miller, 2002). When using this approach in the Latin American context, studies have estimated the percentage of variance in student achievement allocated at the classroom level to be somewhat larger; around 11% in language and 22% in mathematics (Murillo, 2007a; UNESCO, 2008a). Similarly, a review of 17 studies that estimate the magnitude of teacher effects using multilevel covariate adjustment models and gain scores models found that between 7% and 21% of the variance in achievement gain is attributed to teacher effects, and that the size of the estimates depends partially on the subject area being tested and the grade level of the students (Nye, Konstantopoulos and Hedges, 2004).

When implemented, the cross-classified random effects model produces very different estimates of the overall magnitude of teacher effects than do simple covariate adjustment and gain scores models. Raudenbush and Bryk (2002) found that the classroom contribution is estimated to be about 83% of the individual component of the variance of increments to learning in mathematics per year (although part of this variance is due to school to school differences, not specified in their model). Other studies on teacher effectiveness using growth models have also concluded that teachers vary substantially in their effects on individual student learning growth. For example, Rowan et al. (2002) found that the classrooms to which students were assigned in a given year accounted for roughly 60–61% of the variance in students' rates of academic growth in reading achievement, and 52–72% of the variance in students' rates of academic growth in mathematics achievement. Similarly, Palardy (2010) found that the percentage of reading achievement growth between classrooms within schools for kindergarten and first grade was 70%. These estimates are 3 to 10 times the magnitude established in the literature using covariate adjustment and gain scores models. This large discrepancy is likely to be attributable to the better measurement properties of growth curve models.

Also, numerous international studies have suggested greater between-classroom than between-school variance in student achievement and larger teacher effects in primary than in secondary schools for both language and mathematics (Creemers and Reezigt, 1992; Hill and Rowe, 1996; Luyten, 2003; Muijs et al., 2014; Reezigt, Guldemond and Creemers, 1999; Scheerens and Bosker, 1997; Teddlie and Reynolds, 2000).

2.1.5.2 *The Consistency of Teacher Effects*

While research on the consistency of teacher effects across different subjects is scarce, Brophy and Good (1986) cite a study showing a correlation of .70 for adjusted, classroom-level gains across tests of word knowledge, word discrimination, reading, and mathematics. However, Rowan et al. (2002) found only a moderate degree of consistency in classroom effects across reading and mathematics achievement, with correlations ranging from .30 to .47 (depending on the grade level of the classrooms under study) suggesting that a given teacher varies in effectiveness when teaching different academic subjects. These findings, it is worth noting, are comparable to findings on the consistency of school effects across subjects (Scheerens and Bosker, 1997). Campbell et al. (2003), Muijs et al. (2005) and Sammons et al. (2007) also find, in their reviews of the literature, some indication of differentiated effectiveness by curriculum and subject areas.

2.1.5.3 *The Predictors of Teacher Effects*

While EER has demonstrated that a large proportion of the classroom-level variance can be explained by what teachers do in the classroom (Marzano, Pickering and Pollock, 2001; Scheerens and Bosker, 1997; Townsend, 2007), in this section the effect of teacher input variables is discussed, as they are the focus of subsequent analyses.

Several studies have investigated the impact of teacher qualities, such as teaching experience, qualifications, certification and knowledge, as well as the effect of their initial training and working conditions (Darling-Hammond, 2000), providing rather mixed evidence (Darling-Hammond, 2000; Goldhaber and Brewer, 2000; Hanushek, 1997; McCaffrey et al., 2003; Monk, 1994; Wayne and Youngs, 2003; Wilson, Floden and Ferrini-Mundy, 2001). All in all, teacher qualities have been found to have significant but small effects on student achievement (Hanushek, 1989; 1997; Rivkin, Hanushek and Kain, 2005). A review carried out by Greenwald et al. (1996) found that for teacher test scores, the average effect size was *d*=0.12, and for years of experience and postgraduate studies the average effect sizes were less than *d*=0.05. Hattie's review (2009), in turn, found that the overall effect sizes for teacher training and for teacher subject matter knowledge were *d*=0.11 and *d*=0.09, respectively.

Hanushek, Kain, O'Brien and Rivkin (2005) used panel data of students and teachers in Texas to estimate the variation in teacher effects through a value-added approach based on student progress. The authors found a positive relationship between teacher effects and teacher characteristics such as certification, qualifications and teaching experience (with higher gains during the first years of teaching). Similar results were obtained by Clotfelter et al. (2007a), Goldhaber and Brewer (2000) and Clotfelter et al. (2007b), in terms of the relevance of teacher credentials, certification processes and outcomes as predictors of student achievement. Based on data from the state of Kentucky, Kukla-Acevedo's study (2009) found other significant predictors of student achievement, such as teacher's prior achievement, and identifies interactions between teacher-level variables, such as experience, and student-level variables, such as SES and ethnicity.

Rockoff (2004), in turn, used matched student-teacher data, from the state of New Jersey, where both student achievement and teacher data were collected in multiple years. The author used a random-effects meta-analysis approach to measure the variance of teacher fixed effects while taking explicit account of estimation error. His results indicate that teaching experience significantly raises student test scores, particularly in reading.

Finally, the study by Muñoz and Chang (2007) in the state of Kentucky used a multi-level growth curve model and found that teacher experience, education, and race did not predict high school reading achievement growth.

2.1.5.4 *Teacher Cumulative Effects*

Although there is vast research on the issue of how teachers impact students' academic outcomes, less attention has been paid to the continuity of teacher effects measured at different stages of a student's school career (Bressoux and Bianco, 2004; Kyriakides and Creemers, 2008; McCaffrey et al., 2003; Rowan, Correnti and Miller, 2002). This gap in the

EER literature can be attributed to how demanding investigating teacher cumulative effects can be, as it requires analysing high-quality longitudinal student data and applying advanced statistical models. Indeed, apart from research carried out in the United States, studies on the measurement of teacher effects using value-added models that allow the analysis of cumulative effects are scarce, due to the lack of annually administered standardised tests for cohorts of students (O'Donnell and Sargent, 2011).

Raudenbush and Bryk (2002) used a cross-classified model to estimate teacher effects and incorporated a multiple membership component in order to depict teachers' cumulative effects over time. They found that this model fits the data significantly better than the cross-classified model that does not consider the multiple membership of students to teachers. Using a similar method, Kyriakides and Creemers (2008) investigated the long-term effect of schools and teachers in mathematics using longitudinal data from Cypriot students during their first four years of primary school, and concluded that traditional approaches of measuring educational effectiveness tend to overestimate the short-term effects of teachers and student background factors, and underestimate the long-term effects of teachers and schools.

Other studies that have examined the cumulative effects of teachers over time have found significant effects of varying sizes between earlier teachers and subsequent students' academic success (Antoniou, 2012; Hill and Rowe, 1998; Jordan, Mendro and Weerasinghe, 1997; Kyriakides and Creemers, 2008; Pustjens et al., 2007; Rivkin, Hanushek and Kain, 2005; Rowan, Correnti and Miller, 2002; Thum, 2003). In this line of research, Sanders and Rivers (1996) indicated that students with repeated exposure to well-qualified teachers performed up to 50 percentile points better on mathematics tests than those with the same repeated exposure to poorly qualified teachers. Similarly, Tymms et al. (2000) showed that effective classroom experiences in the first years of schooling continue to have a positive influence on students two years later.

Finally, the evidence available is inconsistent with regard to the magnitude of the long-term effects of teachers, and more longitudinal research is needed in this area.

The following conclusions, relevant to the present study, emerge from EER: (1) educational effects are more accurately estimated when achievement growth over time, rather than achievement status, is studied; (2) school effects are larger in emerging economies than in post-industrialised countries; (3) the magnitude of teacher effects tend to exceed school effects; (4) school and teacher effects are not necessarily consistent across subjects and student groups; and, (5) the effects of teachers are likely to accumulate over time. In the following section, the current state of educational effectiveness research in Chile is discussed, and relevant knowledge gaps are identified.

2.1.6 *Educational Effectiveness Research in Chile*

As pointed out by several reviewers, EER has undergone a process of marked internationalisation (Creemers and Kyriakides, 2008; Reynolds et al., 2014). In Latin America and the Caribbean, school effectiveness research can be traced back to the late 1970s and its particular evolution has been described as eminently applied, focused in equity issues and eclectically influenced by a wide range of theoretical currents (Muñoz-Izquierdo, 1996; Murillo, 2007b; UNESCO, 2008a; Velez, Schiefelbein and Valenzuela, 1993). In this region, the study of school and teacher effects has been hampered by the lack of suitable longitudinal data. Thus, very few studies have been designed that could allow for the estimation of these effects on student achievement growth, that is, that have measured student achievement in at least three occasions during their school trajectories and that provide equated achievement scores[10]. As it has been frequently stressed in previous reviews of the Latin-American literature, more longitudinal evidence is needed in order to understand the effect of schools and teachers on children's cognitive growth over time in the region (Murillo, 2007a).

In Chile, the existing educational effectiveness studies have all used cross-sectional data or, at most, two time points longitudinal data, either from the System for Measuring Educational Quality (SIMCE), a national large-scale cross-sectional student assessment programme (e.g., Lara, Mizala and Repetto, 2010; Mizala, Romaguera and Ostoic, 2004;

---

[10] Examples of such research are the GESIS study, a longitudinal research programme focused on elementary school effectiveness in Brazil (Brooke and Bonamino, 2011), and the Young Lives Project (Bourdillon and Boyden, 2014), a long-term international study investigating the changing nature of childhood poverty in Peru and other countries.

Muñoz-Chereau and Thomas, 2015; Troncoso, Pampaka and Olsen, 2015), or from international studies in which the country has participated, such as the Programme for International Student Assessment (PISA) (e.g., Cariola, Cares and Lagos, 2009), the Trends in International Mathematics and Science Study (TIMSS) (e.g., Ramírez, 2006) and the UNESCO comparative studies (e.g., Willms and Somer, 2001). Furthermore, the existent research has mainly focused on the secondary school level and later stages of primary education.

Previous studies on school effects in Chile using raw cross-sectional data have estimated a between-school variance ranging from 30 to 50% (Bellei, 2005; Mizala and Romaguera, 2002; Mizala, Romaguera and Ostoic, 2004; OECD, 2013a; Willms and Somer, 2001). When applying contextualised value-added models, school effects are found to be in the order of 6 to 20% (Muñoz-Chereau and Thomas, 2015, Troncoso, 2015). Also, in line with the international literature, Murillo and Roman (2011) found bigger primary school effects in mathematics than in reading, and in Year 6 than Year 3.

Recently, school effectiveness studies in Chile have started applying more complex value-added models by analysing the variation on student achievement coming from sources other than pupils and schools and/or by incorporating school compositional effects. For example, Mizala and Torche (2012) implemented cross-sectional two-level models for pupil attainment in language and mathematics using SIMCE data. Their analysis of the Chilean school system covered the effects of various pupil- and school-level characteristics, and emphasised the effects of within-school and between-school socioeconomic stratification. The authors also attempted to control for within-school diversity by adding the standard deviation of an indicator of family socioeconomic status (SES) as an explanatory variable at the school level. In another recent study, Manzi et al. (2014) implemented a two-level random coefficients contextualised value-added model controlling for parental qualifications, prior attainment, average school prior attainment, and school SES. They demonstrate how the inclusion of average school prior attainment controls for bias in the estimation of the school effects induced by school selectivity. Also, using contextual value-added models, Muñoz-Chereau and Thomas (2015) and Troncoso (2015) in Chile, and Cervini (2009) in neighbouring Argentina, demonstrate that classroom effects are indeed relevant, as the classroom-level variance is consistently larger than the school-level variance component.

In Chile, as well as in the rest of Latin America, teacher effects and their correlates have rarely been explored (Velez, Schiefelbein and Valenzuela, 1993). As mentioned above, research on the impact of teachers on student achievement is scarce in part due to the difficulty of linking student and teacher data and the lack of longitudinal data suitable for applying teacher value-added models. Indeed, most of the existing studies on teacher effectiveness in Chile have used cross-sectional student achievement data (i.e., Alvarado et al., 2012; Lara, Mizala and Repetto, 2010; León, Manzi and Paredes, 2009; Ortúzar et al., 2009; Ramírez, 2006; Willms and Somer, 2001). Previous research on teacher effects in Chile has been restricted by technical difficulties in modelling relationships between students and their successive classroom settings. Furthermore, most of these studies have not been able to disentangle the teacher contribution from that of the school, nor have they used value-added approaches to estimate teacher effects. Thus, their results are likely to be biased (McCaffrey et al., 2003).

Finally, previous studies on teachers in Chile have often looked at the input teacher variables associated with student achievement. Teacher gender, certification, years of teaching experience and being trained in programmes with subject specialisation and strong practicum components are factors usually related to student outcomes (Lara, Mizala and Repetto, 2010; Ortúzar et al., 2009).

2.1.7 *Model Proposal*

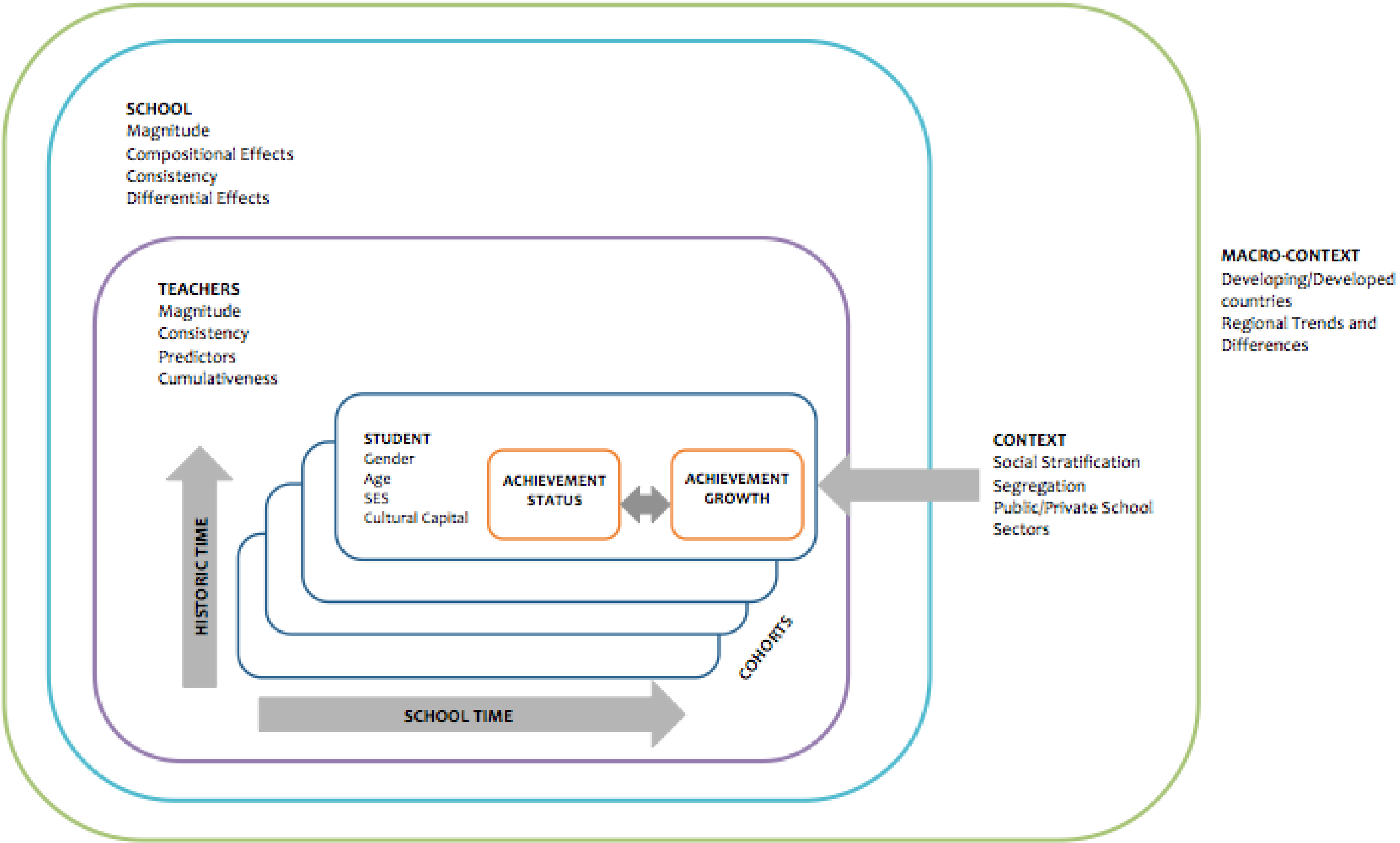
The present study builds upon and contributes to the field of EER by jointly analysing different properties of the effects of schools and teachers on children's cognitive growth over time in Chile, an emergent economy with a socially stratified and segregated schooling system. The project advances the field methodologically by demonstrating the combined use of state-of-the-art techniques, namely: a cohort-sequential design, growth curve approaches and cross-classified multiple membership models. This is the first study in the Chilean context that makes use of annually collected longitudinal data on student achievement linked to extensive student, teacher and school information.

The model in Figure 6 depicts the conceptual framework of this study. This integrated model shows the sources of influence as different levels starting with the macro-context level, which features the distinction between developing and developed countries, as well Latin American trends and within-region differences. Then, the context level refers to the socio-historical events and powers in Chile that have given rise to the educational system and its specific features, such as mixed (public/private) provision, stratification and segregation. The multilevel structure of the model also helps to illustrate the school and teacher influences on the dynamic process of student achievement development. Here, the properties of school and teacher effects to be investigated (i.e., magnitude, predictors, consistency, differentiation and cumulativeness) are depicted. At the student level, the potential influences of the variables gender, age, SES and cultural capital are illustrated. Finally, at the within-student level, the model depicts the relationship between achievement status and growth.

The framework fits the main features of this study, which include a longitudinal design with repeated measures on multiple cohorts. This model expands previous EER models, such as the Dynamic Model (Creemers and Kyriakides, 2008), by putting emphasis on the effect of time. Here, two aspects of time are considered: historic time, on the Y-axis, reflects change in conditions for different cohorts and school time, on the X-axis, depicts change in students' achievement as they move to upper grade levels.

In this chapter, the relevant literature and key concepts for the present study were discussed. The research reviewed provides support for the formation of relevant research aims and the methodological approach to be presented in the following chapter.

FIGURE 6: CONCEPTUAL FRAMEWORK



CHAPTER 3: RESEARCH AIMS AND METHODOLOGY

3.1 AIMS OF THE STUDY

The present study addresses three primary aims, namely:

1. To analyse the achievement trajectories of primary students in Chile, with emphasis on the shape and student-level predictors of achievement growth and their implications for educational equity (Chapter 4).
2. To investigate school effects on student achievement, by analysing their magnitude, compositional effects, consistency across subjects (i.e., language and mathematics) and differential impact across student groups (Chapter 5).
3. To examine teacher effects, with focus on their magnitude, consistency across subjects, teacher-level predictors and cumulative nature (Chapter 6).

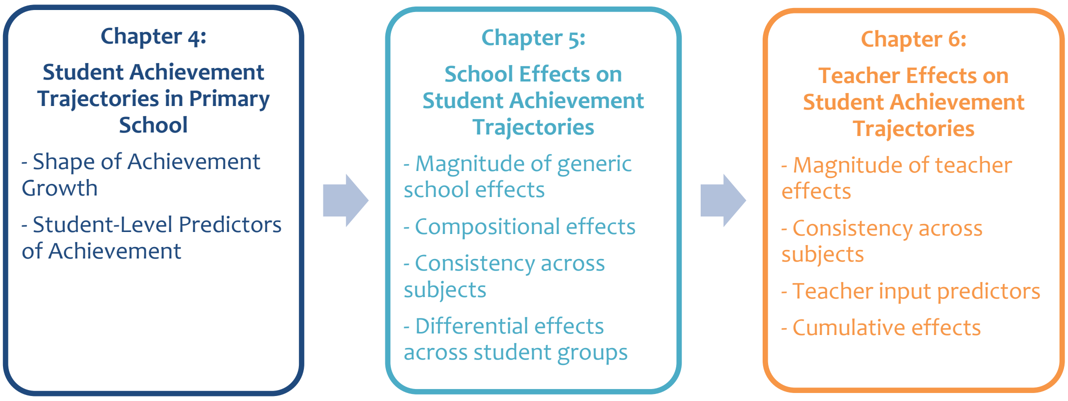
In this chapter, the overarching methodological approach for empirically addressing these research aims is presented. However, in unfolding each one of these aims, more specific research questions and technical issues arise. These are presented and dealt with in greater detail in each corresponding chapter (Chapters 4, 5 and 6).

3.2 RESEARCH DESIGN

The research design of this project involves a sequential analytical progression that builds upon contemporary theoretical models of educational effectiveness, and more specifically, upon the dynamic model of educational effectiveness (Creemers and Kyriakides, 2008). This conceptual model has been introduced in the literature review section of the thesis (Chapter 2). However, it is important to highlight one of the most salient features of this conceptual framework, which refers to the multilevel nature of factors affecting student achievement, with emphasis on the impact of schools and teachers on learning outcomes. This multilevel approach, and several of the different factors identified by the model as aspects that affect educational attainment, define and guide the different phases of the present study. Furthermore, the issues of generic and differential educational effects and of consistency of these effects are also acknowledged and explored in the following chapters.

The first stage focuses on analysing change on student achievement (Chapter 4). The features of models in Chapter 4 provide the basis for further models that estimate individual school and teacher effects and different aspects concerning them (Chapters 5 and 6). This strategy, which involves developing a baseline model from which to start adding different measures of increasing complexity, is what has been defined as bottom-up approach in multilevel analysis (Hox, 2010). The following diagram describes the sequential design of the study:

FIGURE 7: RESEARCH DESIGN



The specific analytical approach for each of these phases is explained in detail in Chapters 4 to 6. In this chapter the focus is on describing the common methodological aspects for the different phases of the study.

3.3 Sources of Secondary Data

For this study, several sets of data have been linked to form a unique database of student, teacher, and school records. The data used in these analyses are derived from two different assessment programmes and three administrative records systems maintained by the Chilean Ministry of Education. It is important to note that the process of selection of sources, and variables within those sources, was informed by both the literature on educational effectiveness and the researcher's familiarity with the available data. A total of 36 data sets, derived from the following sources of secondary data, were obtained, processed and merged together:

- **The SEPA[11] Project.** This student assessment programme, developed by the MIDE-UC Centre of the Pontificia Universidad Católica de Chile, started in 2007. Since 2009 the data sets generated by this project provide standardised measures of student achievement in the subjects of language (Spanish) and mathematics, which are used as the outcome measures of the present study. The SEPA tests are designed to measure student performance on the grade-level competencies specified in the national curriculum standards and are, therefore, criterion-referenced assessments. Both, the language and mathematics tests, consist of 35 multiple-choice items in Grade 3, 40 in Grades 4 to 7, and 50 in Grade 8. The SEPA tests are administered annually, at the end of the school year. The psychometric properties of the tests are presented in Table 28 in Appendix 1. As shown in the table, for each year and grade level considered, the language and mathematics achievement scales present satisfactory estimates of internal consistency (Cronbach's $\alpha > 0.847$). Scores have been both vertically and horizontally equated using Item Response Theory (IRT), which makes scores comparable across both, grade levels and cohorts. The SEPA project was developed with the aim of informing individual schools about their students' overall progress in comparison to that of students in other similar schools and can be characterised as a low-stakes assessment initiative for both students and schools, since results are not publicly available and do not carry consequences, although schools and pupils receive detailed feedback on their performance through confidential reports. These rich longitudinal data have not been used before for research purposes. In the SEPA databases each student is associated to the year(s) and subject in which he/she has been assessed, as well as to his/her school. Furthermore, the national ID number of each student is collected, which makes it possible to confirm the students' grade level and class attended each year against the official records in

[11] Spanish acronym for 'Sistema de Evaluación de Progreso en el Aprendizaje'. In English, 'Learning Progress Evaluation System'.

the MINEDUC School Enrolment Recording System (presented below). This information also allows the linkage of each student with the teacher(s) who taught him/her language and mathematics in each of the years studied, according to the data in the SIGE Recording System (also described below). As a result, any change in results from one test to the next in a particular subject can be linked to the teacher of that subject (assuming no teacher turnover during the year). After carefully considering the quality of these data, it was decided that only the information available for primary school students in the SEPA databases from the years 2010 to 2012 would be used in the present study. This decision was made for two reasons: equated scores were only available for students assessed on those years, and, after merging with other data sources, only the proportions of missing data for students in the primary school level were considered acceptable.

- **The Student Enrolment Recording System (SERS).** At the beginning of each academic year, the Ministry of Education conducts an official survey of students enrolled in the system. These databases allow the identification of the year level and class that the students assessed by the SEPA Project attended each year. They also provide information regarding students' demographic characteristics (i.e., *gender* and *age*) and school-level characteristics (i.e., *school size*).
- **The SIGE[12] Recording System**. These national administrative records are managed by the Ministry of Education. The system annually matches each teacher in the Chilean schooling system with the schools, grades, classes and subjects they taught. This information was extracted for the students in the SEPA sample.
- **The SIMCE[13] Assessment System.** This national assessment programme is carried out by the Ministry of Education and, for each cohort assessed, it includes the entire national population of public and private schools and their students. In the context of this study, these data sets provide individual background data derived from a survey of parents of the students who sat the SIMCE tests. The parent questionnaire provides information on the socioeconomic characteristics of these students, including *family income* and *parental education* (variables used in this

[12] Spanish acronym for 'Sistema de Información General de Estudiantes'. In English, 'General System of Student Information'.
[13] Spanish acronym for 'Sistema Nacional de Evaluación de Calidad Educativa'. In English, 'Educational Quality Measuring System'.

study to create a family socioeconomic status indicator), as well as *number of books at home*. These data were linked to the students assessed by the SEPA Project. The system also collects and publishes relevant information on schools, which are used as covariates in this study and include schools' *sector*, *SES*, *rurality*, *achievement mean* and *achievement standard deviation*. The two later measures are aggregated scores on the SIMCE national standardised reading and mathematics tests, administered by the Ministry of Education to fourth graders. The SIMCE test scores are standardised to have a mean of 250 and a standard deviation of 50[14].

- **The Teacher Census.** These administrative databases provide information on the demographic characteristics (i.e., *gender*), training (i.e., *duration of initial teacher training programme undertaken* and *major in language or mathematics*) and professional experience (i.e., *number of years teaching*) for an important proportion of teachers in both the private and public school sectors. These data were linked to the teachers in our sample.

Appendix 1 provides a more detailed description of the information systems and assessment projects presented above, including aims, sample sizes, and, when relevant, characteristics of the instruments and examples of items.

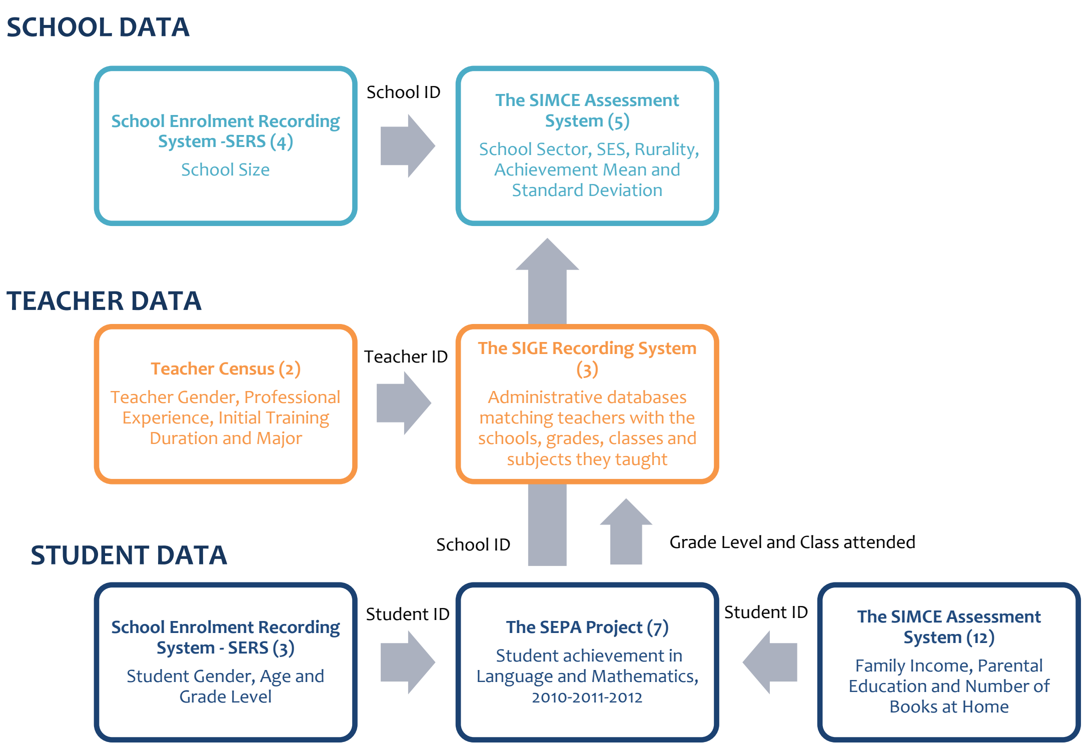
The links between secondary data sources used for each of the chapters (representing each one of them a different research aim of the study) are presented in the following figure. The number of data sets used from each secondary source is shown in parenthesis. While boxes in blue indicate those sources used for analysis in Chapter 4, the analyses on the following chapters use these same linked data but also required adding variables from other sources. As the diagram shows, students, teachers and schools unique identifiers enabled most of the linkages across data sets and over multiple years.

[14] The SIMCE test is a census of students and schools administered annually to pupils of a specified grade level with a schedule that, since 2005, gives the SIMCE test every year to 4th graders and rotates between 8th, and 10th grades and includes parental, teacher, and head teacher questionnaires. However, as the present study looks at the primary school level and, in order to use scores that are comparable across schools, only scores derived from 4th grade testing were considered.

An important link was that between each student's SEPA test score in a specific subject in a given year, and the teacher who taught that particular subject to that student that year. This link was allowed by the grade level and class group identification data available in both, the School Enrolment Recording System and the SIGE Recording System.

Based on the researcher's prior knowledge of many of these data sets, and a review of the newly encountered data elements, it became evident that the quality and availability of the data would have to be considered in depth, as it would substantially affect decisions about the types of analyses that could be conducted. In particular, properties that are often cited in the record linkage literature (Herzog, Scheuren and Winkler, 2007; Winkler, 1995), such as relevance, accuracy, timeliness, accessibility, clarity, comparability, coherence, and completeness, were carefully considered during the process of choosing, requesting, preparing, merging and transforming the data.

FIGURE 8: LINKS BETWEEN SOURCES OF SECONDARY DATA (36 DATA SETS)



Chapter 4: Student Achievement Trajectories and Educational Inequality in Primary School
Chapter 5: School Effects on Student Achievement Trajectories
Chapter 6: Teacher Effects on Student Achievement Trajectories

Some important aspects of the quality of the secondary data used are:

- Student assessment data are available for two subjects, namely language (Spanish) and mathematics.
- Most data sets provide generally complete and reliable information; particularly those that provide school data. However, the teacher and student data sets have some missing or incomplete information.
- The information on student background, derived from the SIMCE Assessment System, was available for most of the cohorts assessed by the SEPA project from 2010 to 2012. Data completeness was only problematic for students who, in 2010, where in Year 2 or lower, and for those who, in 2012 were in Year 8 or higher. Thus, information from four cohorts in primary school was selected in order to use the most complete data available.

The advantages of linking these data sources largely outweigh the disadvantages, especially as this meticulous process of exact matching at the student level provides rich data that allow the modelling of individual performance trajectories, which cannot be accomplished through student data aggregation at the teacher or school level. In the following section the sample resulting from the combination of these data sets is described.

3.4 DATA

3.4.1 *ANALYSIS AND TREATMENT OF MISSING DATA*

After linking the different sources of secondary data, deleting students with implausible grade trajectories, and selecting only the cases belonging to the four cohorts of interest and that had teachers associated to them in both subjects and for each of the three years in which the data were collected (i.e., 2010, 2011 and 2012), the missing data patterns and mechanisms were analysed in depth (results are reported in Appendix 2).

In the language data set, of the 590,817 cells (20,373 cases X 29 variables) in the data matrix, 76,380 of the cells (12.93%) were empty; that is, they had missing values. Similarly, in the mathematics dataset, of the 565,268 cells (19,492 cases X 29 variables) in the data matrix, 70,562 of the cells (12.48%) were empty. Also, on average, variables had missing values on 12 to 13% of the cases (2,634 in language and 2,433 in mathematics) and cases had, on average, missing values on 12 to 13% of the 29 variables considered in the analysis (3.75 in language and 3.62 in mathematics). This confirms that, if listwise deletion had been used, a considerable part of the information in these data sets would have been lost.

The largest proportions of missing data were found in the student-level variables retrieved from the SIMCE Assessment System. *Family income* (19.0% in the language data set and 19.2%), *mother's educational level* (19.0% in language and 19.3% in mathematics) and *father's educational level* (21.7% in language and 21.9% in mathematics), the three variables used for creating a *student socioeconomic status* (SES) indicator, had missing data, as it did the variable *number of books at home* (19.3% in language and 19.6% in mathematics). Also, due to student and school attrition, as well as to the incorporation of new students and schools into the project each year, scores had a large proportion of missing values at each time point. For language test scores, the percentage of missing data were 31.1%, 44.0% and 45.0% in 2010, 2011 and 2012, respectively. Similarly, for mathematics test scores, the percentage of missing data were 32.7%, 42.0% and 42.2% in 2010, 2011 and 2012, respectively. The school-level variables, in turn, presented negligible proportions of missing data. For example, the school variable *achievement mean* presented 0.1% of missing data in both subjects.

Large proportions of missing data and systematic differences between students with complete and incomplete test data are problematic and likely to lead to biased estimates (Bock, Wolfe and Fisher, 1996; McCaffrey et al., 2003; Raudenbush, 2004; Rubin, Stuart and Zanutto, 2004). Also, missing data may reduce statistical power and the precision of calculated statistics. This is particularly the case for variance and covariance estimates (Schafer and Graham, 2002), which are of particular interest in the present study. Despite this, most researchers in the field use complete case analysis, which is only appropriate if the missing data are missing completely at random (MCAR; i.e., missing data is unrelated to the observed data, not missing in a systematic way). This is a strict assumption that most of the time does not hold, and therefore bias is introduced (Enders, 2006).

Missing value analysis helps to address several concerns arising from incomplete data. The analysis of missing data patterns and mechanisms, presented in Appendix 2, indicate that data is not MCAR. As the theorem of ignorability can never be contradicted by the observed data, this is, it is impossible to distinguish empirically between Missing at Random (MAR) and Not Missing at Random (NMAR) response mechanisms (Little and Rubin, 2002), it is assumed that the data is MAR. Thus, it was decided to perform Multiple Imputation (MI) on the data.

MI replaces the missing value of each case with a set of plausible values, using predictors (i.e., auxiliary variables) that are believed to be related to the missing data. This procedure aims to capture the uncertainty about the right value to impute. The multiple imputed data sets are then used in the analysis applying standard statistical methods for complete data, and the results from the imputed data sets are combined. This modern missing data method makes better use of the observed information, increases robustness to non-ignorable missingness and improves estimation precision (Schafer and Graham, 2002).

However, the complex hierarchical structure of the data poses challenges to MI. Research that deals with missing data generally ignore the clustering structure in hierarchical data when performing data imputation, but capacity for MI in two-level structures has recently been developed in statistical packages (van Buuren, 2011). Bayesian MI was performed using Mplus (Muthén and Muthén, 2010). In Mplus Version 7, MI of missing data is generated from a Markov chain Monte Carlo (MCMC) simulation (Asparouhov and Muthén, 2010), a method that was pioneered by Rubin (1987) and Schafer (1997).

Based on recommendations by Rubin (1987) five imputed datasets were generated. Rubin suggests that the efficiency of estimation based on a finite number of imputations, say μ, relative to one based on an infinite number is $(1 + {}^{\lambda}/_{\mu})^{-1}$, where λ is the rate of missing information. Replacing μ with 5 and λ with .13 (a value that approximates the rate of unit-missing data in the sample) gives an estimate of .97 – high enough for the purposes of these analyses.

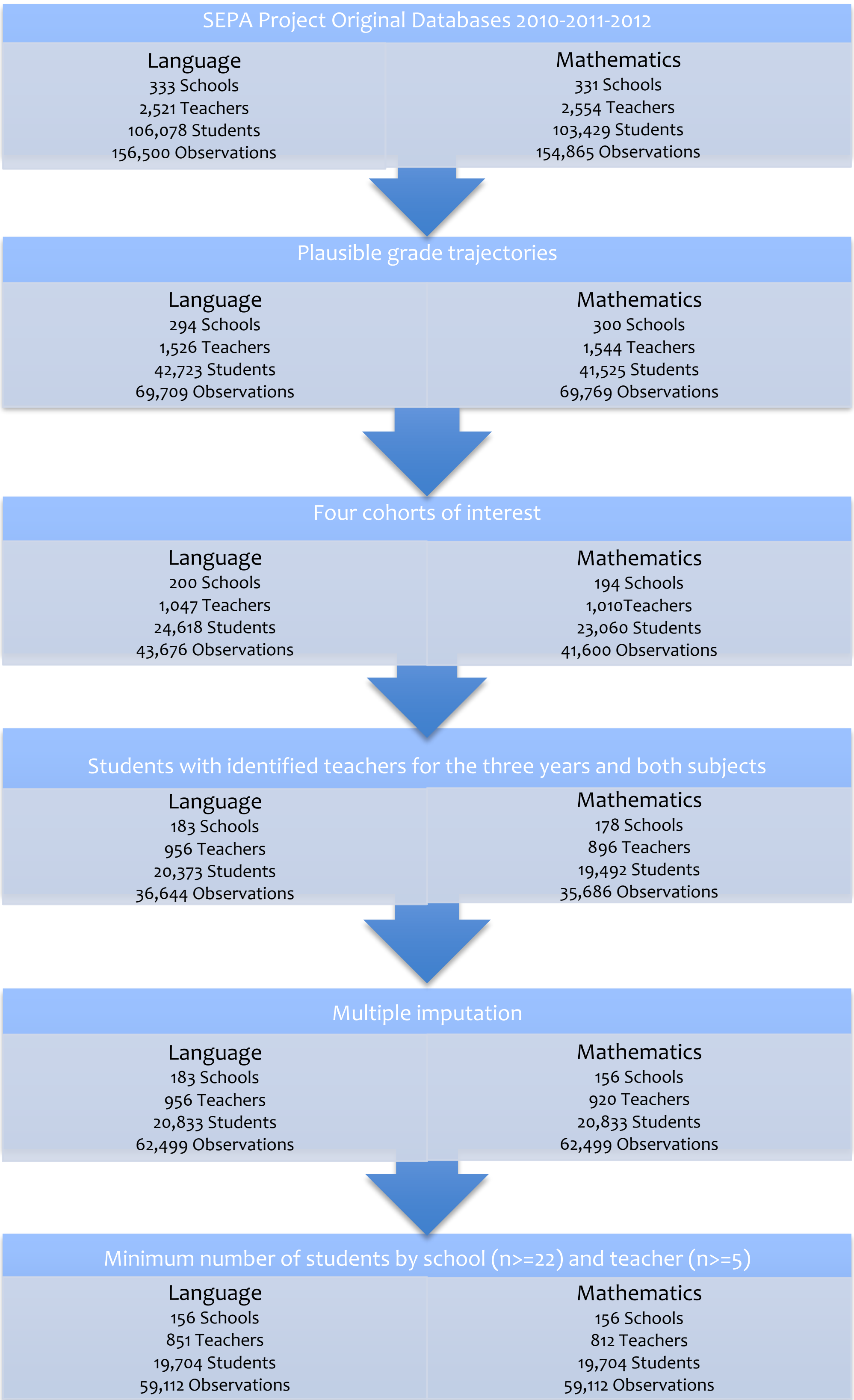
The missing data were imputed after the MCMC sequence had converged. Mplus runs 100 MCMC iterations and then stores the generated missing data values. This process was repeated until the five imputations were stored. These imputed missing data sets are essentially independent draws from the missing data posterior. The missing data was imputed from an unrestricted two-level model and the hierarchical structure was accommodated by means of imputing data with test scores in wide format, students as level 1, and schools as level 2. The language and mathematics databases were linked together so the imputation of data in the language data set would benefit from information on mathematics test scores as auxiliary variables, and vice versa. Finally, all the results obtained from the five multiply imputed data sets were combined using Rubin' rules (1987).

3.4.2 *Sample Size*

The flow chart in Figure 9 shows the methodological decisions in the selection of data and the sample sizes at each step, leading to the final working sample. The development of student achievement in mathematics and language is investigated in four cohorts[15]. These cohorts were in Grade 3 (age 8), 4 (age 9), 5 (age 10) and 6 (age 11) of primary school in 2010[16]. Students who changed school during the period of the data collection were not followed to their new schools as the SEPA Assessment Project is circumscribed to a specific sample of schools. For each student there were three potential occasions of measurement (i.e., years 2010, 2011 and 2012). The measurements were taken at fixed regularly spaced occasions once every year, in November, towards the end of the school year[17].

[15] There were also three consecutive waves of achievement data available for cohorts that in 2010 were in Grade 1 (age 6), 2 (age 7), 7 (age 12) and 8 (age 13) of primary school, and year 1 (age 14) of secondary school. However, the proportion of students from those cohorts that was possible to link to all of the relevant data on student-, teacher- and school-level characteristics ranged between 0% and 35% in language and between 0% and 37% in mathematics. Given the potential bias that these large proportions of missing data could cause to multiple imputation, these cohorts were excluded from the analyses.
[16] Chilean children normally start primary school the year they reach the age of 6.
[17] In Chile, the school year starts in March and finishes in December.

FIGURE 9: FLOW CHART OF SAMPLE SELECTION DECISIONS AND SAMPLE SIZES BY LEVEL



The present study only considers children with plausible grade trajectories in the analysis. Implausible trajectories such as grade acceleration[18] (i.e., jumps of two or more grade levels) were excluded from the sample. Also, students who were retained once or more times during the period analysed were not included in the analyses. Furthermore, analyses were carried out considering only those schools with 22 or more students and teachers with 5 or more students in the years assessed.

Finally, after performing multiple imputation, the language and mathematics samples are balanced, that is, the number of time-point observations is three for each of the students. Figure 9 above shows the final sample size (after multiple imputation was performed) in terms of number of occasions, students, teachers and schools, the classifications in the data. Only the number of teachers is different between the language (N=851) and the mathematics (N=812) samples.

As mentioned above, the data included participants belonging to four different student cohorts, each followed over 3 years. The grade levels in which these cohorts are located each year are presented in Table 2, where cohorts are identified by Roman numerals. Cohort I comprises students who were third graders in 2010, Cohort II those who were fourth graders in 2010, etc. The last cohort, IV, covers students who enter the estimation sample as sixth graders in 2010.

As shown in Table 3, the sample sizes vary by cohort, ranging from n = 4,269 for Cohort 3 to n = 5,782 for Cohort I. The data resembles a 3-year accelerated longitudinal design[19] with four overlapping cohorts, permitting the study of grades 3 to 8. Descriptive statistics of SEPA language and mathematics scores are also summarised in the table by cohort and grade level.

---

[18] Grade acceleration is not a common practice in the Chilean school system; therefore, the presence of records showing this progression might be due to grade miscoding.

[19] Accelerated longitudinal designs are described in greater detail in Chapter 4.

TABLE 2: ESTIMATION SAMPLE USING A THREE-YEAR DATA WINDOW

Grade Level	2010	2011	2012
3	I	-	-
4	II	I	-
5	III	II	I
6	IV	III	II
7	-	IV	III
8	-	-	IV

TABLE 3: DESCRIPTIVE STATISTICS OF LANGUAGE AND MATHEMATICS ACHIEVEMENT BY GRADE LEVEL AND COHORT

	Grade Level					
	3	4	5	6	7	8
Cohort I						
n	5,782	5,782	5,782	-	-	-
M Language (SD)	350.601 (21.342)	377.004 (21.648)	387.978 (19.901)	-	-	-
M Mathematics (SD)	346.168 (20.383)	368.356 (20.948)	374.612 (20.147)	-	-	-
Cohort II						
n	-	5,108	5,108	5,108	-	-
M Language (SD)	-	373.423 (21.159)	393.327 (20.769)	403.865 (19.122)	-	-
M Mathematics (SD)	-	366.066 (20.722)	380.887 (22.146)	397.011 (20.033)	-	-
Cohort III						
n	-	-	4,269	4,269	4,269	-
M Language (SD)	-	-	387.064 (21.585)	406.427 (20.705)	416.093 (19.366)	-
M Mathematics (SD)	-	-	378.354 (20.820)	401.178 (22.206)	419.984 (19.748)	-
Cohort IV						
n	-	-	-	4,545	4,545	4,545
M Language (SD)	-	-	-	403.841 (20.996)	423.353 (22.430)	440.272 (19.430)
M Mathematics (SD)	-	-	-	399.711 (21.284)	424.657 (21.908)	446.473 (20.449)

Note: Dashes indicate that data are not available in that grade for that cohort.

3.4.3 *SAMPLE CHARACTERISTICS*

The geographical location of the participant schools is presented in Figure 10. As the map shows, the sample is highly diverse in terms of school location, although there is a concentration of schools in more populated areas such as the capital city, Santiago, and other main cities in northern and central Chile.

FIGURE 10: GEOGRAPHICAL LOCATION OF PARTICIPANT SCHOOLS

Table 4 presents a series of descriptive statistics that permit comparisons between the population of Chilean students from Grade 3 to Grade 8 of primary school in 2011, those students in the original SEPA project samples and the resulting imputed sample. Most of these school and student variables are used later when modelling student achievement growth and educational effects, as the literature has shown that they can affect student learning.

The SEPA project involves a rather broad range of schools and students. It includes students attending schools that are located in large urban cities and in small rural towns, publicly and privately administered, and from a range of wealth ranging from high to low SES. However, as Table 4 shows, the SEPA samples are somewhat biased in terms of their composition when compared to the relevant student population. While the percentage of students attending rural schools is very similar to those observed in the population (11.2% in the population and 11.8% in the SEPA samples on average), the SEPA project largely over-represents students whose schools are located in the Metropolitan Region[20], with respect to other regions in the country (on average 38.6% of the students attends schools in the Metropolitan Region in the population, compared to 71.2% in the original SEPA samples). Furthermore, while students in the SEPA project samples come from schools with similar average school performance to that of the population of students (less than a fifth of a standard deviation of differences in the SIMCE tests), they are more likely to come from larger schools. Also, the percentage of students attending private non-subsidised and high SES schools are larger in the SEPA project samples than in the population. Finally, the original SEPA samples are representative of the population in terms of student gender but differ somewhat from the population in variables such as number of books at home and parental education, indicating that students participating in the SEPA project typically come from more advantageous home environments.

These differences are likely to be an artefact of both the way in which the project operates and Chile's highly socially stratified education system. The SEPA project is not a census nor is it a survey of schools randomly sampled. Instead, individual schools, or municipal corporations that administrate groups of public schools, decide voluntarily to participate in the project and have to pay for this service. This self-selection process may introduce an equalising force across the sample, in the sense that those head-teachers and municipality administrators who are more confident about their schools' academic performance, have more sophisticated

[20] Chile is divided into 15 regions, which are the country's first-level administrative division. The Metropolitan Region is where the capital city, Santiago, is located.

assessment practices in place, are less averse to external assessment and more motivated about improving their students' academic performance, and can pay for the implementation of this assessment, are more likely to participate in the project. Students who come from higher SES backgrounds and whose families show higher levels of cultural capital are, in turn, more likely to attend those schools.

When comparing the characteristics of the population with those of the final imputed sample, the differences observed with the population are maintained and, in some variables, accentuated. Indeed, despite of the large and diverse group of schools and students comprised in the final sample, it includes only students that progressed without delay through the three years of primary education in which they were or could have been assessed by the SEPA project, and schools and teachers with a minimum number of students (22 and 5, respectively). This filtering process introduces some moderate differences with the original SEPA samples and heightens some of the differences with the population.

Also, it is important to note that the number of students included in the imputed sample represents a small proportion of the relevant student population (0.60% of the students in Grade 3 to Grade 8 of primary schools, in the cohorts and grades of interest).

In conclusion, this research project involves a rather broad range of schools and students. Thus, results obtained from this sample are likely to be more generalisable than studies with more circumscribed samples. However, students who are younger, come from more advantageous home environments and attend high-SES, private and large schools are over-represented in the final analytical sample and those attending low-SES, public and private subsidised schools are underrepresented. The implications of this for the study are a potential reduction in variance at the school and teacher levels, which would mean that the estimates of school and teacher effects might appear lower than they actually are. Thus, the differences between the characteristics of the population and the composition of the final sample demand caution with regard to the interpretation and generalisation of findings.

TABLE 4: COMPARISON OF THE FINAL SAMPLE WITH THE ORIGINAL SEPA SAMPLES AND THE POPULATION

	Source	Population 2011[1]	Original SEPA Project samples 2010-2011-2012[2]		Final sample 2010-2011-2012[3]
			Language	Mathematics	
		(N=1,513,277)[4]	(N=106,078)	(N=103,429)	(N=19,704)
SCHOOL VARIABLES					
Proportion of students in rural schools	SERS-STUDENT	0.112	0.117	0.119	0.092
Proportion of students in public schools	SERS-STUDENT	0.434	0.394	0.402	0.377
Proportion of students in private subsidised schools	SERS-STUDENT	0.495	0.399	0.398	0.401
Proportion of students in private non-subsidised schools	SERS-STUDENT	0.071	0.207	0.200	0.222
School SES (1 to 5)	SIMCE-SCHOOL	2.840	3.013	3.000	3.120
School size (number of students in primary school)	SERS-STUDENT	522.990	675.006	662.041	881.327
School overall reading performance	SIMCE-SCHOOL	267.012	271.543	270.790	271.031
School overall mathematics performance	SIMCE-SCHOOL	259.179	263.961	263.206	263.418
School reading performance SD	SIMCE-SCHOOL	44.679	43.986	44.132	44.009
School mathematics performance SD	SIMCE-SCHOOL	42.381	44.007	44.125	44.141
STUDENT VARIABLES					
Proportion of females	SERS-STUDENT	0.490	0.482	0.479	0.499
Age (in December 2010)	SERS-STUDENT	10.950	11.175	11.320	10.664
Number of books at home (1 to 5)	SIMCE-STUDENT	2.902	3.148	3.140	3.093
Family income (1 to 13)	SIMCE-STUDENT	4.199	5.547	5.506	5.456
Mother's level of education (1 to 20)	SIMCE-STUDENT	12.601	13.339	13.306	13.243
Father's level of education (1 to 20)	SIMCE-STUDENT	12.636	13.569	13.539	13.412

Sources: The SIMCE Assessment System (SIMCE, Grade 4, Primary level, 2011 database) and School Enrolment Recording System (SERS, Primary level, 2011 database).

[1] Variables were calculated using students from Grade 3, primary school to Grade 8, primary school, except for school SES and school overall performance, that are calculated using only for students in Grade 4, primary school as only this data is available at the population level.

[2] It includes students from Grade 1, primary school to Grade 3, secondary school.

[3] It includes students from Grade 3, primary school to Grade 8, primary school.

[4] According to SERS student data (primary level, grades 3 to 8, 2011 database).

3.5 METHOD

3.5.1 MULTILEVEL MODELS

In order to address the research aims of the study and appropriately account for the structure of the data, several multilevel models are fitted and progressively developed. The multilevel models implemented in this study correspond to Contextualised Value-Added Models (CVAM). Table 5 summarises the main features of the models introduced in Chapters 4 to 6. The equations and parameters of interest are fully described in the corresponding chapters. In what follows, the general analytical approach for each chapter is briefly presented.

3.5.1.1 *ACCELERATED GROWTH MODELS (CHAPTERS 4 TO 6)*

As it will be explained in the following chapter, the complex structure of the longitudinal data in this study presents many challenges for statistical modelling. The models introduced in Chapters 4 to 6 are implemented to accommodate both, multiple measures of the same student (via growth curve modelling), and multiple cohorts of students (using accelerated longitudinal modelling). From a substantive point of view, by developing three-level models and adding relevant covariates, these models allow the analysis of student achievement trajectories, the identification of significant student-level predictors of achievement status and growth leading to achievement gaps, and the estimation of the magnitude and consistency of school effects.

Also, as discussed in the literature review chapter, previous research has suggested that schools might have differential effects on individual student learning. In this study, differential effects of schools across student groups are explored using accelerated growth models and adding school-level random slopes for the student covariates gender and socioeconomic status. This provides a means of modelling school effects that vary across groups of students and that are indicative of within-school equity.

3.5.1.2 *CROSS-CLASSIFIED AND MULTIPLE MEMBERSHIP ACCELERATED GROWTH MODELS (CHAPTER 6).*

The proportion of variance of primary student achievement growth at the teacher level is estimated using several features of the models developed in Chapters 4 and 5. However, as some students change teachers across years, student outcomes do not follow the traditional nested design of hierarchical models and an alternative model specification is necessary, namely, a cross-classified random effects model. As discussed in the literature review in Chapter 2, the model for estimating teacher effects from repeated measurements of student achievement when students encounter multiple teachers over time was proposed by Raudenbush (1993) and later extended by Rowan, Correnti and Miller (2002) to include schools as a third level in which both students and teachers are nested. The model, also denominated crossed random effects growth model (Palardy, 2010), conceives teacher effects on student growth as 'deflections', either positive or negative, from each child's specific growth trajectory. In addition to analysing the magnitude of teacher effects, the correlation between teachers' value added scores in language and mathematics, for the subsample of teachers that taught both subjects, is examined in order to assess whether teachers have consistent effects on students' achievement growth across academic subjects.

This study's data set also allows the investigation of the effect of some teacher characteristics on student achievement growth. In this section of the thesis, teacher input variables that explain the variation in teacher effects are identified by using a cross-classified accelerated growth model and adding the fixed effects of teachers' years of experience, gender, subject specialisation and initial teacher training duration.

Finally, the hypothesis of cumulative effects of teachers is also tested by comparing the fit of the model that assumes that teachers from prior years make no contributions to current achievement growth with the model that assumes that previous teachers' effects persist undiminished in future years. The latter model specifies the teacher effect on student achievement growth in a given year as the joint effect of all of the previous teachers the student had for that subject, during the period considered in the study, with the contribution of each teacher being assigned an equal weight[21]. This is done by adding a multiple-membership component to the cross-classified growth model. In Chile, students are frequently assigned to different teachers each year and prior research has not been able to incorporate the potential cumulative effects of teachers. The complexity of the structure of this formal framework of analysis is expected to match the system being studied to a larger extent than models used in previous research.

[21] While persistence parameters can be specified and not treated as fixed, this is computationally challenging (Lockwood et al. 2007) and will not be explored in the present study.

TABLE 5: COMPARISON FEATURES OF MULTILEVEL MODELS BY CHAPTER

Model	Number of Levels / Classifications	Growth Model with Linear and Quadratic Terms of Time	Accelerated Longitudinal Design	Cross-Classified Model	Multiple Membership Model	Fixed Effects	Random Effects
Chapter 4							
Model 0	2						Random intercept and slope at the student-level
Model 1	2	X	X			Cohort effects	Random intercept and slope at the student-level
Model 2	3	X	X			Cohort effects	Random intercepts and slopes at the student and school levels
Model 3	3	X	X			Cohort effects + student-level variables	Random intercepts and slopes at the student and school levels
Chapter 5							
Model 4	3	X	X			Cohort effects + student-level variables + school-level variables	Random intercepts and slopes at the student and school levels
Model 5	3	X	X			Cohort effects + student-level variables + school-level variables	Random intercepts and slopes at the student and school levels and random slope for student gender/SES at the school level
Chapter 6							
Model 6	4	X	X	X		Cohort effects + student-level variables + school-level variables	Random intercepts and slopes at the student, school and teacher levels
Model 7	4	X	X	X		Cohort effects + student-level variables + school-level variables + teacher-level variables	Random intercepts and slopes at student, school and teacher levels
Model 8	4	X	X	X	X	Cohort effects + student-level variables + school-level variables	Random intercepts and slopes at student, school and teacher levels

3.5.2 *ESTIMATION*

Models such as the ones fitted in this study, with complex error structures and potentially small variance components, have seen only limited application in the literature, mainly due to software limitations and estimation difficulties (particularly for maximum likelihood–based estimators applied to large datasets) (Palardy, 2010; Raudenbush, 2008). In these circumstances, estimation within a Bayesian framework, applying Markov chain Monte Carlo (MCMC) algorithms has been preferred over maximum likelihood estimators (Browne and Draper, 2006; Goldstein, 2011).

3.5.2.1 *BAYESIAN ANALYSIS*

Bayesian analysis is a statistical paradigm that describes unknown parameters using probability statements. Under this approach, a parameter is summarised by an entire distribution of values instead of one fixed value as in classical frequentist analysis (Gelman et al., 2004). This posterior distribution, constructed from a prior distribution about a parameter and a likelihood model providing information about the parameter based on observed data, can be approximated using MCMC methods.

Unique features of Bayesian analysis include an ability to incorporate prior information in the analysis, an intuitive interpretation of credible intervals as fixed ranges to which a parameter is known to belong with a pre-specified probability, and the ability to assign an actual probability to any hypothesis of interest (Jackman, 2009). This approach has found important applications in multilevel analysis (Draper, 2008; Seltzer, Wong and Bryk, 1996).

Under MCMC methods, the posterior distribution is used to form various summaries for the model parameters, including point estimates such as posterior means, medians, percentiles, and interval estimates known as credible intervals. As chains start at particular starting values it will generally take a while for the chains to converge and sample from the actual posterior distribution. The period when the chains are settling down is normally called the burn-in period and these iterations are omitted from the sample from which summaries are constructed.

In this study, estimation was performed using the Bayesian estimation via MCMC methods implemented in the software MLwiN (Browne, 2012). All models were run for a burn-in of 10,000 iterations followed by a monitoring chain of 100,000 iterations[22]. The means and standard deviations of the sampled parameters from the monitoring period were used as parameter estimates and standard errors while the 2.5th and 97.5th quantiles of these chains provided Bayesian 95% credible intervals, analogous to 95% confidence intervals. Also, vague prior distributions were specified (the default in MLwiN).

Informal visual assessments of the parameter chains and standard MCMC convergence diagnostics suggest that the samplers were run for sufficiently long. Appendix 3 presents the estimates and traces change as the estimation proceeded through the 100,000 iterations, for each of the parameters in Model 8, the most complex of the models fitted in this study, as well as the convergence diagnostics for the random-effects parameters in the model. The plots of the estimated Monte Carlo Standard Error (MCSE) of the posterior estimate of the mean against the number of iterations were also examined. The MCSE is an indication of the accuracy of the mean estimate ($MCSE = SD/\sqrt{n}$, where SD is the adjusted standard deviation from the chain of values, and n is the number of iterations). These graphs were used to calculate how long to run the chains to achieve a mean estimate with a particular desired MCSE.

Also, two contrasting accuracy diagnostics were inspected: the Raftery-Lewis and the Brooks-Draper diagnostics. The Raftery-Lewis diagnostic (Raftery and Lewis, 1992) is a diagnostic based on a particular quantile of the distribution and is used to evaluate the length of Markov chain required to estimate a particular quantile to a given accuracy. In MLwiN the diagnostic is calculated for the two quantiles (the defaults are the 2.5% and 97.5% quantiles) that will form a central interval estimate. The maximum estimated chain length suggested for the school and teacher random parameters in Model 8 was 13,074, so having run the chain for 100,000 iterations this diagnostic has been satisfied. The Brooks-Draper diagnostic is based on the mean of the distribution. It is used to estimate the length of Markov chain required to produce a mean estimate to k significant figures

[22] The default number of iterations for the burn-in period and the monitoring period are 500 and 5,000, respectively.

with a given accuracy. In Appendix 3, it is possible to see that to quote the school and teacher random parameter estimates with the desired accuracy requires the chains to be run for a maximum of 5,312 iterations, so this diagnostic is also satisfied for all of the parameters in Model 8 and for the rest of the less complex models.

The MLwiN software was operated via the Stata command *runmlwin* (Leckie and Charlton, 2012). The *runmlwin* command can fit the full range of multilevel models available in MLwiN seamlessly from within Stata using both the IGLS and MCMC algorithms and provides full control over all aspects of model specification and estimation. This command works by writing, sending and then running an MLwiN macro file for the specified model in MLwiN and then returns, stores and displays the model results in Stata, which, in turn, offers several options for hypothesis testing, model comparison, graphics and other post-estimation commands for interpretation and analysis. Carrying out analyses using MLwiN's point-and-click interface is usually inefficient and error prone, particularly when fitting a large number of models. This issue is overcome by using runmlwin that permits conducting the analyses in a documentable and reproducible manner. Also, by using the Stata command *mi estimate* in combination with *runmlwin,* results are computed for the five multiple imputed datasets and combined using the combination rules by Rubin (1987).

Finally, depending on the model to be fitted, estimates obtained through the quasi-likelihood methods implemented in MLwiN, from the immediately less complex model or from the same model but for the other subject, were used as starting values.

3.5.2.2 Model Fit

When models are fitted using MCMC algorithms, the Bayesian Deviance Information Criterion (DIC) is recommended to compare model fit. DIC is a 'badness' of fit indicator developed specifically for use on complex hierarchical models and Bayesian estimators, and allows the comparison of non-nested models and penalises for model complexity (Hamaker et al., 2011). The DIC is a generalisation of Akaike's Information Criterion (AIC), a likelihood-based measure for comparing non-nested models (Gelman et al., 2004; Spiegelhalter et al., 2002). As with the AIC, the DIC combines goodness of fit with model complexity (the number of parameters), so that DIC values for different models can be compared directly. Lower values reflect superior models and differences in DIC values of more than 5 units between two models are regarded as strong evidence in favour of the model with the smaller DIC (Lunn et al., 2012).

3.6 DATA MANAGEMENT

The original file formats of the different data sets that were merged to create the unique databases used in this study varied widely across the different secondary sources but the most common were .exe, .sav, .csv. These files were imported, transformed and stored into SPSS. Metadata (i.e., data about data) accompanying and describing these data files, provided by the organisations managing the secondary sources, were saved in electronic form. Metadata for the final database were developed and included in each data file in the form of variable and value labels.

The processes of record linkage through exact matching, data cleaning, missing data analysis and calculation of descriptive statistics were performed in SPSS. Multiple imputation of missing data was performed using the Mplus software as it allows for multiple imputation of hierarchical data. Finally, as mentioned above, multilevel modelling was conducted using the software package MLwiN via Stata, because of the former's flexibility for specifying models with up to five levels and cross-classified data structures, the possibilities of estimating models via Markov chain Monte Carlo (MCMC) methods, and to benefit from the latter's graphics capability and programming interface. Annotated syntax files were created in each of these statistical packages to record and reproduce analyses.

3.7 ETHICAL CONSIDERATIONS

This research follows the guidelines of both the British Educational Research Association (BERA)[23] and the Central University Research Ethics Committee (CUREC)[24]. Approval for

[23] Guidelines available in: www.bera.ac.uk/publications/guidelines/

the study was granted by the Departmental Research Ethics Committee (DREC). All of the conducted and proposed statistical analyses are based on secondary sources of data, collected and provided by the MIDE-UC Assessment Centre of the Pontificia Universidad Católica and the Ministry of Education of Chile. This study uses de-identified data that have been anonymised by professionals of the Ministry of Education, following the indications of the researcher. Thus, the project meets the appropriate ethical standards set by the Department of Education with regards to the use prior-anonymised data. Further measures were taken to obtain permission from both the MIDE-UC Assessment Centre of the Pontificia Universidad Católica and the Chilean Ministry of Education, which also consented to the use of the data.

[24] Further information in:
https://weblearn.ox.ac.uk/portal/hierarchy/socsci/education/hd/page/research_ethics.

CHAPTER 4: STUDENT ACHIEVEMENT TRAJECTORIES IN PRIMARY SCHOOL

4.1 INTRODUCTION

In this chapter, growth models of language and mathematics achievement measurements (level 1) nested within students (level 2), nested, in turn, within schools (level 3), are estimated to characterize individual achievement growth trajectories and examine the effect of various socioeconomic and demographic variables on achievement growth.

The analyses proceed through three stages: Firstly, the adequacy of curvilinear growth models to each cohort in the data is assessed. Then, overlapping cohorts of data collected on four cohorts over three years are linked to approximate change in achievement between Year 3 and Year 8 of primary school. Finally, the relationship of these trajectories with time invariant covariates at the student level are assessed in order to explore academic achievement gaps related to gender, age, socioeconomic status and cultural capital.

4.2 RESEARCH QUESTIONS

As presented in section 3.1, the first aim of this study is to analyse the achievement trajectories of primary students in Chile, with emphasis on establishing the shape and identifying student-level predictors of achievement growth and the way they influence educational equity. This aim unfolds into the following research questions, to be answered in this chapter:

1. What is the level and change of performance (in language and mathematics) for students in primary school?
2. How large are the differences between students in their achievement and in their changes in achievement over time?

3. What student-level characteristics account for the variation in student growth trajectories over time?

The first research question is about average student achievement at different grades and refers to fixed effects. The second question addresses random effects, namely, differences between students. Finally, the third question refers to the fixed effect of different student covariates in student achievement status and growth rate. As it was mentioned in the previous chapter, these questions are addressed using a multilevel growth curve approach.

4.3 Method of Analysis

4.3.1 Data

In this section the basic growth patterns found in the observed data are described. Figure 11 shows actual language and mathematics growth trajectories for 100 randomly selected students in the sample. These trajectories seem fairly monotonic (increasing with grade), but nonlinear. Additionally, observed growth trajectories anticipate substantive individual variation, both, in achievement levels and growth rates, as achievement levels and growth slopes seem to differ among students.

As mentioned in the previous chapter, our data included participants belonging to four different student cohorts. Indeed, the data resembles a 3-year accelerated longitudinal design[25] having four overlapping cohorts, permitting the study of grades 3 to 8. Based on the descriptive statistics presented in section 3.4.2, Figure 12 shows the mean at each grade for each cohort and its 95% confidence interval. These means are connected by lines to approximate the mean growth tendencies. This analysis provides a first indication of cohort differences. As we can see, adjacent cohort means are somewhat similar, and the confidence intervals of the cohorts overlap in some of the time points. Also, both language and mathematics proficiency show upward growth. Overall, in language there is a tendency for rapid increase during the first years of primary school. This rate seems

[25] Accelerated longitudinal designs are defined later in this chapter.

to decrease somewhat in the final years. Mathematics shows a similar trend although it appears to present more steady growth rates over the years.

The curvilinear shape of these trajectories is consistent with past research in a wide range of national contexts (e.g., Brooke and Bonamino, 2011; Caro et al., 2009; Guldemond and Bosker, 2009; Verhaeghe, Van Damme and Knipprath, 2011). However, this evidence does not support an entirely developmental explanation because cohort and grade level are correlated and may interact. Thus a more detailed analysis is required.

FIGURE 11: ACHIEVEMENT TRAJECTORIES FOR 100 STUDENTS

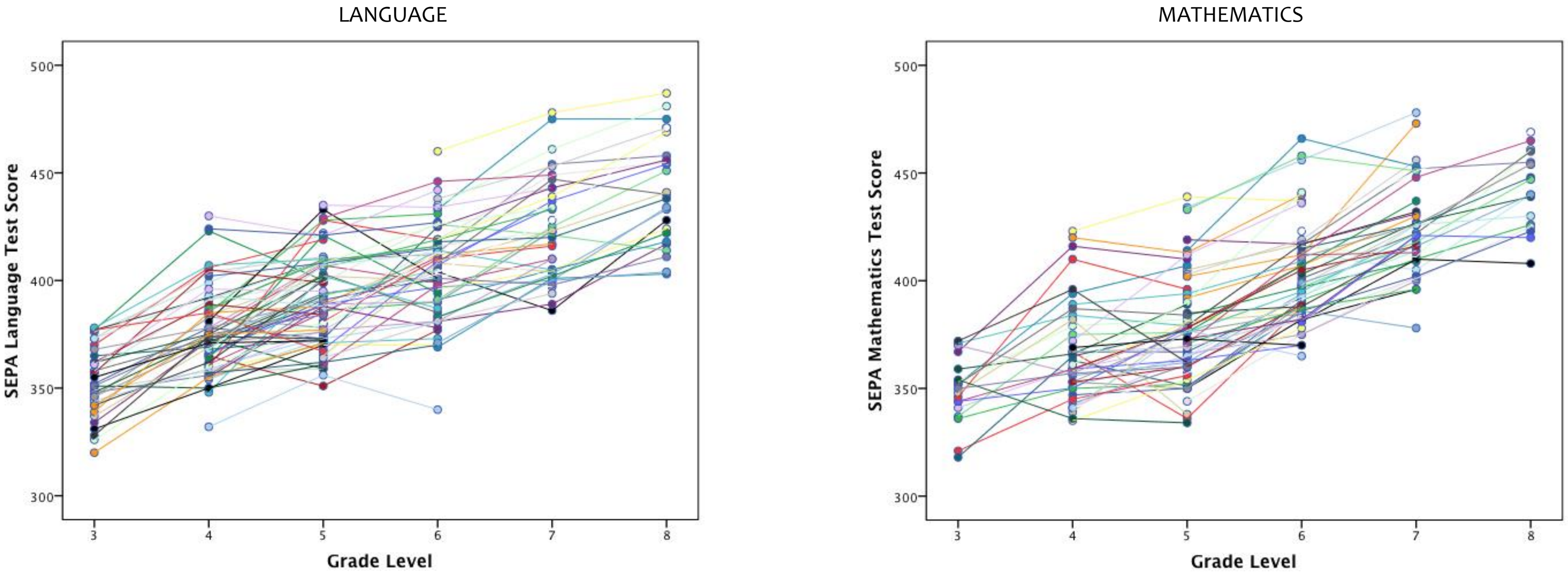
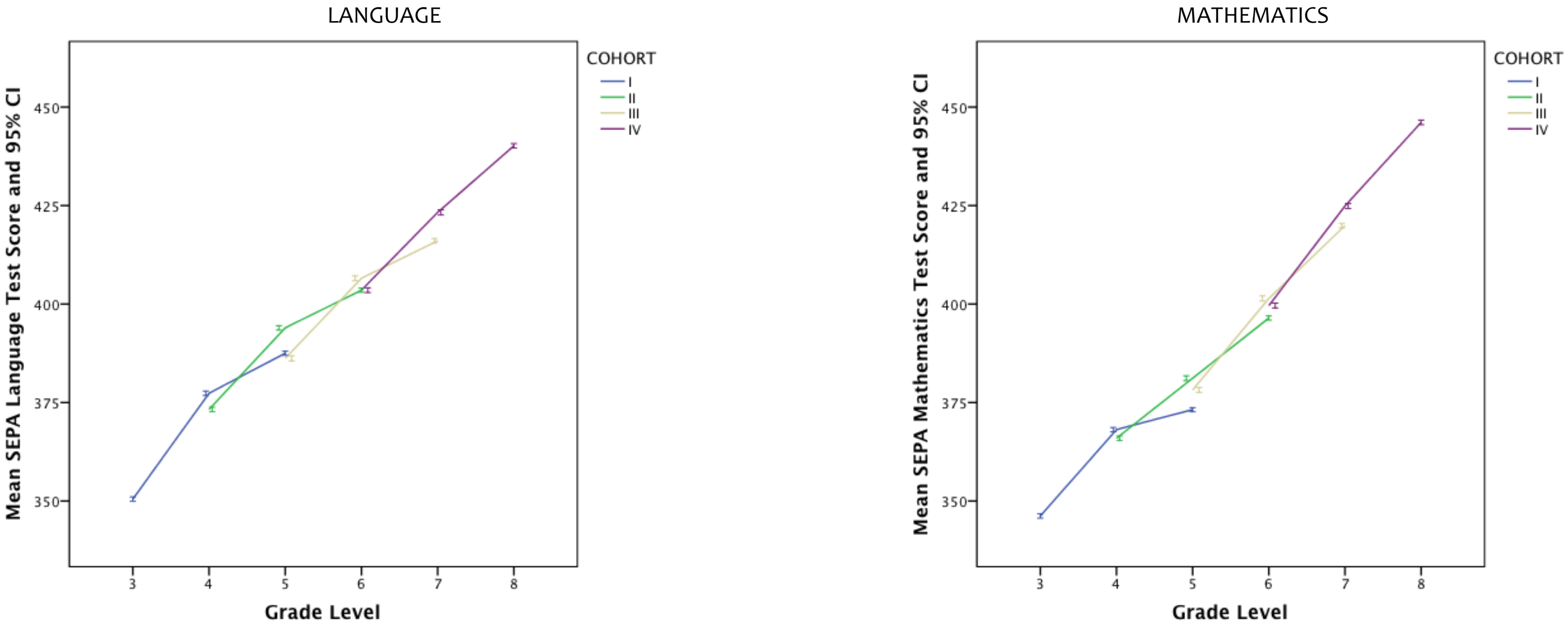


FIGURE 12: SEPA TEST SCORE BY GRADE LEVEL AND COHORT



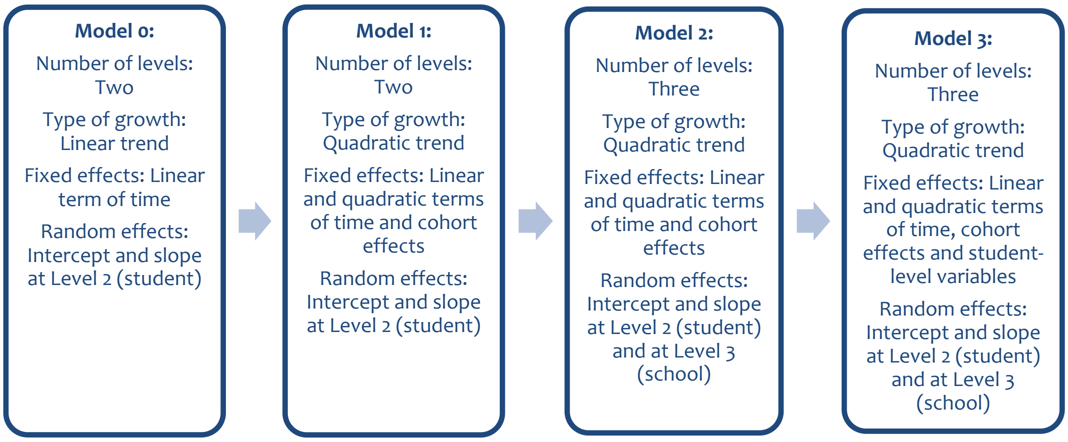
4.3.2 *MODELS*

In this chapter, students' growth in both mathematics and language achievement in primary schools and the student-level characteristics that affect these trajectories are studied. To do so a growth model approach is used in which different trajectories over time for the students are modelled.

Under this approach, the longitudinal data are analysed using multilevel models in which level-one units, measurement occasions (*t*), are nested within level-two units individual subjects (*i*). In this fixed occasion repeated-measures design, '*t*' stands for the time variable. The specification of the fixed part of the model is discussed by de Fraine et al. (2005). The authors compare the multivariate approach (different time points are modelled as dummy-variables) and the growth curve analysis (data are modelled as a function of a time variable) regarding the latter as an elegant and parsimonious way of modelling longitudinal data, suitable for designs such as the one in the present study in which a time dimension underlies the different measurements, the outcome variable is measured on the same scale at every occasion and the number of measurement occasions is three or more.

Figure 13 shows the progression of models explored in this chapter. The analyses are organised in four models. The first step, as in any multilevel model, is the decomposition of the variance in the outcome of interest into its between- and within-group parts. This is done in Models 0, 1 and 2. Model 3 is the within-school model, which estimates how student characteristics affect test scores within schools. In this final model, the influences of student gender, age, family SES and number of books at home on test scores are modelled.

FIGURE 13: GROWTH MODELS IN CHAPTER 4



4.3.2.1 *MODEL 0. A TWO-LEVEL LINEAR GROWTH MODEL*

Three time points provide sufficient data for estimating two random parameters. Therefore, a linear specification that has only a random intercept and a random slope is adopted. Several alternative notations have been proposed in the literature for formalising multilevel models. In the following equations, the notation introduced by Raudenbush and Bryk (2002) is used. Later, in Chapter 8, this specification is extended to multiple subscript notation, introduced by Rasbash and Browne (2001), to allow the description of multilevel models with combinations of hierarchical, crossed and multiple membership components.

EQUATION 1

$$\begin{gathered} \boldsymbol{Y_{ti} = \beta_{0i} + \beta_{1i} t_{ti} + e_{ti},} \\ \boldsymbol{\beta_{0i} = \beta_0 + u_{0i}} \\ \boldsymbol{\beta_{1i} = \beta_1 + u_{1i}} \\ \begin{pmatrix} \boldsymbol{u_{0i}} \\ \boldsymbol{u_{1i}} \end{pmatrix} \sim \boldsymbol{N} \left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\sigma^2_{u0}} & \\ \boldsymbol{\sigma_{u0u1}} & \boldsymbol{\sigma^2_{u1}} \end{pmatrix} \right] \\ \boldsymbol{e_{ti} \sim N(0, \sigma^2_e)}. \end{gathered}$$

Equation 1 specifies the following multilevel model: at the first level, student achievement is described by a linear function of time and the second level describes the variability in the individual growth curves. The fixed part of the model describes the average growth trajectory by means of two parameters: the average intercept ($\boldsymbol{\beta_0}$) and an average linear growth parameter ($\boldsymbol{\beta_1}$). In addition, random effects were introduced for the initial status and growth rates to allow these parameters to vary among students. Thus, the trajectory of each individual is described by two parameters: a student specific intercept ($\boldsymbol{\beta_{0i}}$) and a student specific linear growth parameter ($\boldsymbol{\beta_{1i}}$). The student specific residuals ($\boldsymbol{u_{0i}}, \boldsymbol{u_{1i}}$) are assumed to be generated by a bivariate normal distribution with average zero, variances $\boldsymbol{\sigma^2_{u0}}$ and $\boldsymbol{\sigma^2_{u1}}$ and covariance $\boldsymbol{\sigma_{u0u1}}$ (Raudenbush and Bryk, 2002; Snijders and Bosker, 2012). The error at level 1 ($\boldsymbol{e_{ti}}$) is the error at time *t* for the *i*th individual. These within person residuals are assumed to be mutually independent and normally distributed with mean zero and constant variance $\boldsymbol{\sigma^2_e}$.

The parameter $\boldsymbol{\beta_{0i}}$ denotes the achievement score of student *i* at time *t*=0. Therefore, it is preferable that the zero score for the time variable has an interpretable meaning. Indeed, in repeated measures analysis, careful thought must be given to the coding of the time variable as changing the time point to which the intercept refers alters the meaning of the intercept in growth curve models, affects the correlation between intercept and linear slope, changes the variance of the intercept and has important implications for the interpretation of the models (Anumendem et al., 2013; Biesanz et al., 2004; Singer and Willett, 2003).

Thus, to interpret the correlation between the intercept and the slope, the zero point must have a strong substantive meaning. In this time-structured design, 'grade level' is the time metric chosen. The data are ordered by data collection occasion, with difference in age at the first measurement occasion treated later as a student-level predictor. In the models, the time variable 'grade level' is scaled with the starting point of the three grades observed for each cohort coded as zero, this is, with the intercept as initial status.

Model 0 (i.e., the baseline model) is introduced to determine the variance at the individual level without any explanatory variable. This model is later extended by adding an extra level (i.e., the school level) and by modelling the growth parameters (initial status and linear growth) as functions of student and school characteristics.

4.3.2.2 *Model 1. A two-level growth model with a quadratic effect of time*

Researchers using multilevel analysis to depict growth mostly opt for polynomial growth models, to be able to model nonlinear growth (Guldemond and Bosker, 2009). The linearity of change on student is tested formally in the coming model. To do so, Equation 2 was elaborated to include a quadratic term. Here, a linear specification that has only two parameters, intercept and slope, and includes a curvilinear slope that is fixed across students is adopted.

EQUATION 2

$$\begin{gathered}
\boldsymbol{Y_{ti} = \beta_{0i} + \beta_{1i} t_{ti} + \beta_2 t_{ti}^2 + e_{ti}}, \\
\boldsymbol{\beta_{0i} = \beta_0 + u_{0i}} \\
\boldsymbol{\beta_{1i} = \beta_1 + u_{1i}} \\
\boldsymbol{\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix} \right]} \\
\boldsymbol{e_{ti} \sim N(0, \sigma_e^2)}.
\end{gathered}$$

At the first level, student achievement is described by a linear and a quadratic function of time and the second level describes the variability in the individual growth curves. The fixed part of the model describes the average growth trajectory by means of three parameters: the average intercept ($\boldsymbol{\beta_0}$), an average linear growth parameter ($\boldsymbol{\beta_1}$) and an average quadratic growth parameter ($\boldsymbol{\beta_2}$). The specification of the random part of the model remains as in Equation 1[26].

4.3.2.3 *MODEL 2. A THREE-LEVEL GROWTH MODEL WITH A QUADRATIC EFFECT OF TIME*

Model 2, depicted in Equation 3, is a three-level model consisting of schools, students within schools, and time points within students. At the first level, students' achievement is again described as a quadratic function of time. The second level describes the variability in individual growth while the third level described differences between schools both in terms of achievement status and growth rates.

EQUATION 3

$$\begin{gathered}
\boldsymbol{Y_{tij} = \beta_{0ij} + \beta_{1ij} t_{tij} + \beta_2 t_{tij}^2 + e_{tij}}, \\
\boldsymbol{\beta_{0ij} = \beta_{00j} + u_{0ij}} \\
\boldsymbol{\beta_{1ij} = \beta_{10j} + u_{1ij}}
\end{gathered}$$

[26] The trajectory of each individual could also have been described by a third parameter: a student specific random quadratic growth parameter. However, models with this specification would fit the data perfectly and therefore would not reach convergence. Thus, a quadratic random parameter is not implemented.

$$\beta_{00j} = \beta_{000} + u_{00j}$$
$$\beta_{10j} = \beta_{001} + u_{01j}$$
$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]$$
$$\begin{pmatrix} u_{00j} \\ u_{01j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u00}^2 & \\ \sigma_{u00u01} & \sigma_{u01}^2 \end{pmatrix}\right]$$
$$e_{tij} \sim N(0, \sigma_e^2).$$

4.3.2.4 *MODEL 3. A THREE-LEVEL GROWTH MODEL WITH A QUADRATIC EFFECT OF TIME AND STUDENT-LEVEL PREDICTORS*

Model 3 is fitted to study differences in achievement status and growth related to students' demographic characteristic and social background. The specification of the model is as follows:

EQUATION 4

$$Y_{tij} = \beta_{0ij} + \beta_{1ij} t_{tij} + \beta_2 t_{tij}^2 + e_{tij},$$
$$\beta_{0ij} = \beta_{00j} + \beta_{010} Female_{ij} + \beta_{020} Age_{ij} + \beta_{030} SES_{ij} + \beta_{040} Books_{ij} + u_{0ij}$$
$$\beta_{1ij} = \beta_{10j} + \beta_{110} Female_{ij} + \beta_{120} Age_{ij} + \beta_{130} SES_{ij} + \beta_{140} Books_{ij} + u_{1ij}$$
$$\beta_{00j} = \beta_{000} + u_{00j}$$
$$\beta_{10j} = \beta_{001} + u_{01j}$$
$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]$$
$$\begin{pmatrix} u_{00j} \\ u_{01j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u00}^2 & \\ \sigma_{u00u01} & \sigma_{u01}^2 \end{pmatrix}\right]$$
$$e_{tij} \sim N(0, \sigma_e^2).$$

The four covariates introduced into the model are student characteristics frequently used in models of educational achievement. *Female* is a dichotomous variable that distinguishes boys (0) from girls (1). *Age* refers to student age, calculated in years and months, as in December of 2010 and cohort-mean centered. *SES* is a family socioeconomic status index constructed as a composite of variables indicating parental education and family wealth, following traditional operationalisations of the SES concept (Buchmann, 2002; Hauser, 1994; Yang and Gustafsson, 2004), although indicators of parental occupational status, a common component of SES indices, were not available in the secondary data used. More specifically, the SES index was obtained from a factor analysis of mother's education, father's education and family monthly income, and was standardised to have a mean of zero and a standard deviation of unity. Other operationalisations of SES use the maximum educational level of either parent to reduce the amount of missing information in SES (e.g., OECD, 2009). However, as suggested by Caro and Cortés (2012), the chosen SES model, based on three separate items, yielded greater reliability than the one with two items using the maximum strategy (Cronbach's α = 0.876 vs. 0.827) and produced a greater number of possible item combinations and therefore also of SES scores. Finally, *Number of books at home (books)*, a proxy variable for cultural capital and the value of scholarly culture, was reported by parents and categorised in five values (1 = None, 2 = Less than 10 books, 3 = between 10 and 50 books, 4= between 51 and 100 books and 5 = More than 100 books). The four student-level covariates entered in Model 3 are, by definition, time-invariant. Appendix 4 shows the operational definition and descriptive statistics of these variables.

When estimating the models, continuous covariates at the student level were grand-mean centered and dummy-coded variables were left uncentered. Given that the study's research questions focus on analysing variance across students, schools, and later teachers, the analyses need to control for independent variables across the entire sample. Grand-mean centering accomplishes this objective (Kreft, De Leew and Aiken, 1995; Raudenbush and Bryk, 2002). Also, separate growth models were estimated for each academic subject (language and mathematics).

When analysing the models, the direction and size of the effect of each of the predictors of growth in students' achievement are considered. Analysing how much of the variation both at the student and at the school level are explained by these variables is also of interest. Therefore, the proportion of variance explained and goodness of fit indicators are compared across models.

4.3.3 *The Accelerated Growth Model*

The structure of the data in this study fits into what has been called a cohort-sequential design, or an accelerated longitudinal design. This design takes place when different cohorts of subjects are followed for a relatively short period of time, and then a curve is modelled across the entire span of the data (Tonry et al., 1991). The accelerated longitudinal research design links adjacent segments of limited longitudinal data from different cohorts to create a common long-term developmental trend or growth curve. Thus, researchers can approximate a long-term longitudinal study by conducting several simultaneous short-term longitudinal studies of different cohorts (Duncan and Duncan, 2012). This technique for accelerated collection of longitudinal information was first introduced by Bell (1953) and has recently gained popularity as a method for dealing with longitudinal design problems such as time constrain, subject attrition and the cost of multiple and continual subject assessments (Duncan and Duncan, 2012). An advantage of this approach is that by studying several cohorts, rather than one, confidence in the generalisability of the results can be increased. Different authors have demonstrated that the accelerated design is an efficient method for examining longitudinal data (e.g., Duncan, Duncan and Hops, 1996; Raudenbush and Chan, 1993). Duncan and Duncan (2012) provide a review on research applications of accelerated longitudinal designs, concluding that they have often been used within the field of developmental psychology to study a variety of phenomena, such as youth substance use, family and individual processes, intellectual abilities and health-related behaviours. However, this method has seldom been applied in EER. Indeed, only four educational studies have been found that use this approach (i.e., Biancarosa, Bryk and Dexter, 2010; De Haan et al., 2012; Erikson and Rudolphi, 2010; Zvoch and Stevens, 2006). The present study seeks to advance the field methodologically by demonstrating the application of accelerated longitudinal models in the context of the study of educational inequalities and effectiveness.

Figure 12 in section 4.3.1 depicts the cohort-sequential design of the data. There are four grade cohorts of children who in 2010, the first wave of data collection considered in this study, are in Grade 3, 4, 5 and 6, respectively. The data collection took 3 years, with data collected from the same children yearly. Thus, the four cohorts provide information of progression across three different consecutive grade levels: 3-4-5, 4-5-6, 5-6-7 and 6-7-8. In the final sample, there is a grade level range of 3 to 8, and a growth curve will be fitted across six years, although the actual data collection took only three years. Thus, with an accelerated design, the growth curve is estimated on a combination of cross-sectional and longitudinal information (Hox, 2010).

However, an important issue to be tested is whether the different cohorts are comparable, this is to say, whether there are cohort effects. Such cohort effects may arise because of demographic differences between cohorts and because of effects of history. An analysis pooled from multiple cohorts ignoring their demographic differences would present a distorted picture of grade-related change. If the data collection contains a sufficient number of measurement occasions, cohort effects can be tested by fitting four separate linear growth curves and assessing whether these are equal in the four cohorts (Raudenbush and Chan, 1993). This method, used in the present study, examines whether unspecified effects of demography and history have given rise to cohort differences in grade-outcome trajectories. When the grade-outcome trajectories do not vary by cohort, they may be said to converge.

Missing data in accelerated longitudinal designs is considered to be missing at random (MAR) as missingness is usually planned and under the researcher's control (Duncan and Duncan, 2012). In this study, the data structure resembles an accelerated longitudinal design but missing data might be generated from mechanisms other that the research design, such as the different schools that make up each cohort of students in the study. Therefore, testing for cohort effects becomes even more relevant. Indeed, in the context of this study, the Grade X Cohort interaction effect (usually referred in the literature as Age X Cohort interaction effect, due to sampling based on age being more common in this type of design) represents a threat to valid inference and such an effect would be present if, for example, the rate of increase in achievement differed for the different cohorts because the cohorts differed in background. Given that the different cohorts are observed at different grade intervals, these differences in rates of increase could masquerade as developmental effects (Raudenbush and Chan, 1993). Thus, grade-cohort interaction analyses test the generalisability of within cohort inferences.

Raudenbush and Chan (1993) implement a method for testing cohort effects on age-outcome curves using hierarchical models, with the aim of bringing two overlapping cohorts into the model. Later Miyazaki and Raudenbush (2000) extend this approach by illustrating a procedure that incorporates multiple cohorts into the model to test convergence. To test this, the average status and average rate of change of two adjacent cohorts at the grade of overlap are compared; that is, at grade 4, 5, 6 and 7. If these are very similar it could be assumed that the trajectories of the two cohorts coincide and that a single trajectory is typical of development from grades 3 to 8.

The approach implemented by Miyazaki and Raudenbush (2000), which focuses on the plausibility of believing that the estimated plots for the four cohorts vary by chance from a single underlying developmental continuum, is applied. The idea is to compare two models: a more complex model under which each cohort has its own trajectory and a simpler model according to which all cohorts follow a common trajectory. The authors recommend a three-step procedure for testing convergence of cohort-specific trajectories. First, a cohort-based hierarchical model (the *full* model) is estimated. As in Model 1, at the first level of the model (t), each person's observed development is conceived as a quadratic function of grade level plus random error. At the second level of the model (i), the individual intercept and linear growth rate coefficients are assumed to vary as a function of cohort plus person-specific random effects. Thus separate mean trajectories are estimated for each cohort. Inspection of the data in the previous sections suggests that the mean trajectory has a quadratic shape. Therefore, to test convergence, the full model based on Model 1 is estimated but, in level 2, a distinct average trajectory for each cohort is estimated. Thus, at Level 2, person-specific change parameters depend on cohort membership (Equation 5).

EQUATION 5

$$\boldsymbol{Y_{ti} = \beta_{0i} + \beta_{1i}t_{ti} + \beta_2 t_{ti}^2 + e_{ti}},$$

$$\boldsymbol{\beta_{0i} = \beta_0 + \sum_{c=1}^{3} \beta_{0c} d_{ci} + u_{0i}}$$

$$\boldsymbol{\beta_{1i} = \beta_1 + \sum_{c=1}^{3} \beta_{1c} d_{ci} + u_{1i}},$$

$$\boldsymbol{\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]}$$

$$\boldsymbol{e_{ti} \sim N(0, \sigma_e^2)}.$$

where $\boldsymbol{d_{ci}}$ = 1 if person $\boldsymbol{i}$ belongs to Cohort (c+1) for c = 1,2,3; $\boldsymbol{d_{ci}}$ = 0 if not (here Cohort 1 is the reference group).

Secondly, a common model for all cohorts (the *reduced* model) is estimated (Equation 6). The competing reduced model does not include cohort effects. Instead, individual trajectories are viewed as varying randomly around a single mean-grade trajectory. This model is written as:

EQUATION 6

$$\boldsymbol{Y_{ti} = \beta_{0i} + \beta_{1i}t_{ti} + \beta_2 t_{ti}^2 + \beta_{3i}t\_cohort_{ti} + \beta_4 t\_cohort_{ti}^2 + e_{ti}},$$

where t_cohort_{ti} is a linear component of the function of grade, centered (i.e., coding the zero-point in time) around the cohort's starting grade (i.e., 3 for Cohort 1, 4 for Cohort 2, 5 for Cohort 3 and 6 for Cohort 4) and $t\_cohort_{ti}^2$ is a quadratic component of the function grade centered around the cohort's starting grade. The definition of the other parameters in level 1 of this model corresponds to that of Model 1 (Equation 2). The inclusion of t_{ti}, t_cohort_{ti} and their squares in the same model serves the purpose of ensuring that both the full and the reduced models have the same covariance structure. Collinearity is avoided by using the following specification at level 2 of the model:

$$\boldsymbol{\beta_{0i} = \beta_0 + u_{0i}}$$
$$\boldsymbol{\beta_{1i} = \beta_1}$$
$$\boldsymbol{\beta_{3i} = u_{3i}},$$

where $\boldsymbol{u_{3i}}$ is the random effect of person $\boldsymbol{i}$ on the rate of increase at the grade starting point of the cohort. The random effect associated with t is restricted to zero. In addition, fixed effects for t_{ti} and t_{ti}^2 but not for t_cohort_{ti} and $t\_cohort_{ti}^2$ are estimated.

Finally, since Monte Carlo Marcov Chain (MCMC) estimation method is employed for fitting both the full and reduced models, the goodness of fit of the two models is compared by means of the Deviance Information Criterion (DIC). The smaller the value of DIC, the better the fit to the data. Therefore, if the reduced model presents a better fit compared to the full model, then it can be assumed that all cohorts share the same mean trajectory.

Assembling the components of Model 3 (Equation 4) and the full model produces the final model depicted in Equation 7.

EQUATION 7

$$Y_{tij} = \beta_{0ij} + \beta_{1ij}t_{tij} + \beta_2 t_{tij}^2 + e_{tij},$$

$$\beta_{0ij} = \beta_{00j} + \beta_{010}Cohort2_{ij} + \beta_{020}Cohort3_{ij} + \beta_{030}Cohort4_{ij} + \beta_{040}Female_{ij} + \beta_{050}SES_{ij} + \beta_{060}Books_{ij} + \beta_{070}Age_{ij} + u_{0ij}$$

$$\beta_{1ij} = \beta_{10j} + \beta_{110}Cohort2_{ij} + \beta_{120}Cohort3_{ij} + \beta_{130}Cohort4_{ij} + \beta_{140}Female_{ij} + \beta_{150}SES_{ij} + \beta_{160}Books_{ij} + \beta_{170}Age_{ij} + u_{1ij}$$

$$\beta_{00j} = \beta_{000} + u_{00j}$$

$$\beta_{10j} = \beta_{001} + u_{01j}$$

$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]$$

$$\begin{pmatrix} u_{00j} \\ u_{01j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u00}^2 & \\ \sigma_{u00u01} & \sigma_{u01}^2 \end{pmatrix}\right]$$

$$e_{tij} \sim N(0, \sigma_e^2).$$

Table 6 summarizes the interpretation for the parameters in the models presented above that are of primary interest to research questions 1 and 2[27]. Finally, as explained in Chapter 3, the data on which the analyses were performed were obtained through multiple imputation of missing data.

[27] Research question 3, in turn, refers to more readily interpretable estimates of fixed effects of student-level variables.

TABLE 6: INTERPRETATION OF KEY ESTIMATES

Variable	Variable Description
Intercept	
$\boldsymbol{\beta_{00j}}$	Baseline initial achievement status for students in the starting point of the three grades observed for Cohort 1
$\boldsymbol{\sigma^2_{u0}}$	Variability in initial achievement status between students nested within schools
$\boldsymbol{\sigma^2_{u00}}$	Variability in initial achievement status between schools
Yearly growth parameter	
$\boldsymbol{\beta_{10j}}$	Baseline learning rate during the year in the starting point of the three grades observed for Cohort 1
$\boldsymbol{\sigma^2_{u1}}$	Variability in learning rates among students within schools
$\boldsymbol{\sigma^2_{u01}}$	Variability in school effects on student learning during the baseline period
Yearly quadratic growth parameter	
$\boldsymbol{\beta_{2}}$	Learning rate acceleration during the year in the starting point of the three grades observed
Adjustment for cohort-specific intercept	
$\boldsymbol{\beta_{010}}$	Average adjustment to initial achievement status for Cohort 2
$\boldsymbol{\beta_{020}}$	Average adjustment to initial achievement status for Cohort 3
$\boldsymbol{\beta_{030}}$	Average adjustment to initial achievement status for Cohort 4
Adjustment for cohort-specific growth rates	
$\boldsymbol{\beta_{110}}$	Average adjustment to learning rate for Cohort 2
$\boldsymbol{\beta_{120}}$	Average adjustment to learning rate for Cohort 3
$\boldsymbol{\beta_{130}}$	Average adjustment to learning rate for Cohort 4

Note: When other controls are included in the models, these parameters should be interpreted as above plus the adjustment for the other variables in the model.

4.4 Results

The results section of this chapter is structured as follows. First, student change in language and mathematics achievement is modelled by cohort. Second, a joint growth trajectory across cohorts is estimated and the variance components of the models are analysed. Finally, after controlling for the variance attributable to the school level, the effects of different student-level predictors are examined.

4.4.1 Modelling Change by Cohort

To specify the appropriate shape of the mean and individual change functions, the data from the four cohorts are first analysed separately, using two growth curve models for individual change: one that assumes linear growth (Model 0, Equation 1) and another that tests for quadratic fixed effects (Model 1, Equation 2). For both models, it is expected that the average rate of increase is positive during the primary school years but with substantial variation across students.

Table 7 displays the posterior means and standard deviations (analogous to frequentist parameter estimates and standard errors) of the fixed and random-part parameters for Model 0, the time-nested-within-student random coefficient growth model for language and mathematics, by cohort. This model allows the shape of students' growth trajectories to be identified and determines whether the initial intercept and random grade slope vary across individuals. Fixed effects results show that, for all cohorts in both subjects, there is a clear linear trend of time averaged over students.

According to Model 1, there is a tendency for the achievement score of each student to change at different rates, but the degree of acceleration or deceleration of this rate is common for all students within each cohort[28]. Results in Table 8 indicate that mean achievement growth follows a non-linear (quadratic) trend that applies to all of the cohorts in both subjects, except for Cohort 2 in mathematics. The significant effects for both the linear and quadratic terms for most of the cohorts suggest that both should be retained in the analysis, and the negative effects of the latter indicate that, on average, individuals' growth rates slow down slightly over time (this is found for all cohorts

[28] It could also be that rates of acceleration or deceleration vary across subjects, however, given the number of time points available, a model with quadratic random effects would fit the data perfectly and therefore would not converge.

except for Cohort 2 in mathematics, which does not present a deceleration, nor a significant acceleration, of the growth rate). Thus, it is concluded that both linear and quadratic terms are useful in describing students' growth trajectories. Results presented in Table 8 are also depicted in Figures 14 and 15.

TABLE 7: MODEL 0 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS, BY COHORT

MODEL 0 LANGUAGE								
	COHORT 1		**COHORT 2**		COHORT 3		**COHORT 4**	
	M (SE)	95% CI	**M (SE)**	**95% CI**	M (SE)	95% CI	**M (SE)**	**95% CI**
Fixed Part								
Intercept	353.167 (0.303)	352.571 - 353.763	**374.987 (0.330)**	**374.327 - 375.647**	388.684 (0.604)	387.279 - 390.088	**404.287 (0.394)**	**403.481 - 405.094**
Time	18.689 (0.298)	17.946 - 19.432	**15.222 (0.165)**	**14.871 - 15.572**	14.513 (0.295)	13.792 - 15.234	**18.214 (0.196)**	**17.780 - 18.647**
Random Part								
Random Level 2 (Student)								
Intercept	329.046 (13.879)	298.012 - 360.080	**334.356 (10.539)**	**313.208 - 355.505**	340.729 (12.052)	316.200- 365.258	**345.955 (14.472)**	**314.879 - 377.030**
Time/Intercept	-16.326 (5.504)	-29.798 - -2.855	**-26.487 (3.254)**	**-32.939 - -20.035**	-25.303 (6.000)	-39.530 - -11.076	**-19.233 (8.084)**	**-39.264 - 0.797**
Time	1.871 (0.621)	0.586 - 3.157	**7.028 (2.819)**	**0.886 - 13.169**	3.565 (2.139)	-0.817 - 7.946	**4.476 (2.486)**	**-0.548 - 9.499**
Random Level 1 (Occasion)								
Intercept	153.376 (5.822)	138.598 - 168.154	**126.990 (6.106)**	**111.844 - 142.136**	132.425 (4.166)	123.237 - 141.614	**126.229 (3.066)**	**120.107 - 132.351**
Units: Students	5,782		**5,108**		4,269		**4,545**	
Units: Occasions	17,346		**15,324**		12,807		**13,635**	
DIC	141,925		**123,110**		103,135		**109,111**	
pD	4,992		**4,614**		3,757		**4,298**	

MODEL 0 MATHEMATICS								
	COHORT 1		**COHORT 2**		COHORT 3		**COHORT 4**	
	M (SE)	95% CI	**M (SE)**	**95% CI**	M (SE)	95% CI	**M (SE)**	**95% CI**
Fixed Part								
Intercept	348.818 (0.423)	347.893 - 349.743	**365.852 (0.369)**	**365.089 - 366.616**	379.025 (0.533)	377.827 - 380.223	**400.248 (0.446)**	**399.297 - 401.198**
Time	14.223 (0.724)	12.249 - 16.197	**15.473 (0.258)**	**14.841 - 16.104**	20.813 (0.238)	20.251 - 21.376	**23.380 (0.257)**	**22.754 - 24.005**
Random Part								
Random Level 2 (Student)								
Intercept	290.994 (10.095)	269.757 - 312.231	**324.371 (10.580)**	**302.508 - 346.234**	344.298 (14.525)	312.563 - 376.033	**351.042 (11.135)**	**328.693 - 373.391**
Time/Intercept	-1.727 (2.485)	-6.959 - 3.504	**-4.321 (3.122)**	**-11.464 - 2.823**	-11.283 (4.413)	-21.612 - -0.954	**-9.027 (3.510)**	**-16.470 - -1.584**
Time	0.236 (0.124)	-0.008 - 0.480	**0.321 (0.177)**	**-0.026 - 0.669**	1.123 (0.574)	-0.043 - 2.288	**1.362 (1.003)**	**-0.622 - 3.345**
Random Level 1 (Occasion)								
Intercept	148.552 (6.899)	130.493 - 166.611	**124.549 (2.138)**	**120.195 - 128.902**	116.293 (2.724)	110.421 - 122.165	**115.969 (2.198)**	**111.579 - 120.359**
Units: Students	5,782		**5,108**		4,269		**4,545**	
Units: Occasions	17,346		**15,324**		12,807		**13,635**	
DIC	141,230		**122,086**		100,864		**107,722**	
pD	4,937		**4,526**		3,897		**4,177**	

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions and student levels. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

TABLE 8: MODEL 1 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS, BY COHORT

	COHORT 1		COHORT 2		COHORT 3		COHORT 4	
	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI
Fixed Part								
Intercept	350.603 (0.341)	349.915 - 351.291	373.426 (0.406)	372.579 - 374.272	387.065 (0.729)	385.307 - 388.824	403.840 (0.457)	402.872 - 404.809
Time	34.117 (0.953)	31.746 - 36.487	24.588 (1.244)	21.363 - 27.813	24.211 (1.327)	20.790 - 27.632	20.808 (0.877)	18.724 - 22.892
Time^2	-7.714 (0.581)	-9.219 - -6.210	-4.684 (0.654)	-6.398 - -2.969	-4.848 (0.552)	-6.243 - -3.454	-1.296 (0.377)	-2.166 - -0.427
Random Part								
Random Level 2 (Student)								
Intercept	347.309 (13.267)	318.374 - 376.245	347.786 (10.239)	327.356 - 368.215	356.120 (12.099)	331.686 - 380.555	347.883 (14.502)	316.469 - 379.297
Time/Intercept	-25.026 (5.496)	-37.809 - -12.244	-34.881 (3.444)	-41.753 - -28.009	-34.673 (6.247)	-49.129 - -20.216	-20.486 (8.200)	-41.255 - 0.283
Time	10.707 (3.020)	3.922 - 17.491	15.601 (2.622)	9.972 - 21.231	12.926 (2.618)	7.497 - 18.355	5.566 (2.484)	0.136 - 10.996
Random Level 1 (Occasion)								
Intercept	124.495 (4.200)	114.865 - 134.124	111.071 (3.055)	104.632 - 117.509	115.189 (3.146)	108.761 - 121.616	124.600 (2.553)	119.577 - 129.624
Units: Students	5,782		5,108		4,269		4,545	
Units: Occasions	17,346		15,324		12,807		13,635	
DIC	138,771		121,330		101,860		108,988	
pD	5,962		5,565		4,426		4,408	

	COHORT 1		COHORT 2		COHORT 3		COHORT 4	
	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI
Fixed Part								
Intercept	346.169 (0.431)	345.218 - 347.121	366.070 (0.400)	365.234 - 366.905	378.355 (0.527)	377.190 - 379.519	399.710 (0.440)	398.786 - 400.634
Time	30.153 (1.779)	25.376 - 34.931	14.168 (0.833)	12.181 - 16.155	24.835 (0.639)	23.444 - 26.226	26.511 (0.981)	24.091 - 28.931
Time^2	-7.966 (1.090)	-10.931 - -5.000	0.652 (0.465)	-0.496 - 1.800	-2.010 (0.308)	-2.680 - -1.339	-1.565 (0.503)	-2.823 - -0.307
Random Part								
Random Level 2 (Student)								
Intercept	302.091 (11.825)	276.234 - 327.949	323.804 (10.998)	301.039 - 346.569	347.133 (15.228)	313.622 - 380.644	353.474 (10.911)	331.660 - 375.288
Time/Intercept	-5.123 (4.325)	-15.127 - 4.881	-4.178 (3.758)	-12.647 - 4.290	-12.796 (4.858)	-24.070 - -1.522	-10.432 (3.542)	-17.953 - -2.910
Time	4.427 (3.290)	-3.837 - 12.692	0.717 (0.808)	-0.895 - 2.330	2.352 (1.287)	-0.287 - 4.991	2.339 (1.345)	-0.445 - 5.123
Random Level 1 (Occasion)								
Intercept	123.019 (3.688)	114.526 - 131.511	123.980 (2.552)	118.634 - 129.325	113.764 (3.186)	106.744 - 120.785	114.176 (2.156)	109.917 - 118.434
Units: Students	5,782		5,108		4,269		4,545	
Units: Occasions	17,346		15,324		12,807		13,635	
DIC	137,950		122,084		100,596		107,509	
pD	5,891		4,546		4,046		4,301	

FIGURE 14: GROWTH CURVES FOR LANGUAGE BASED ON MODEL 1, BY COHORT

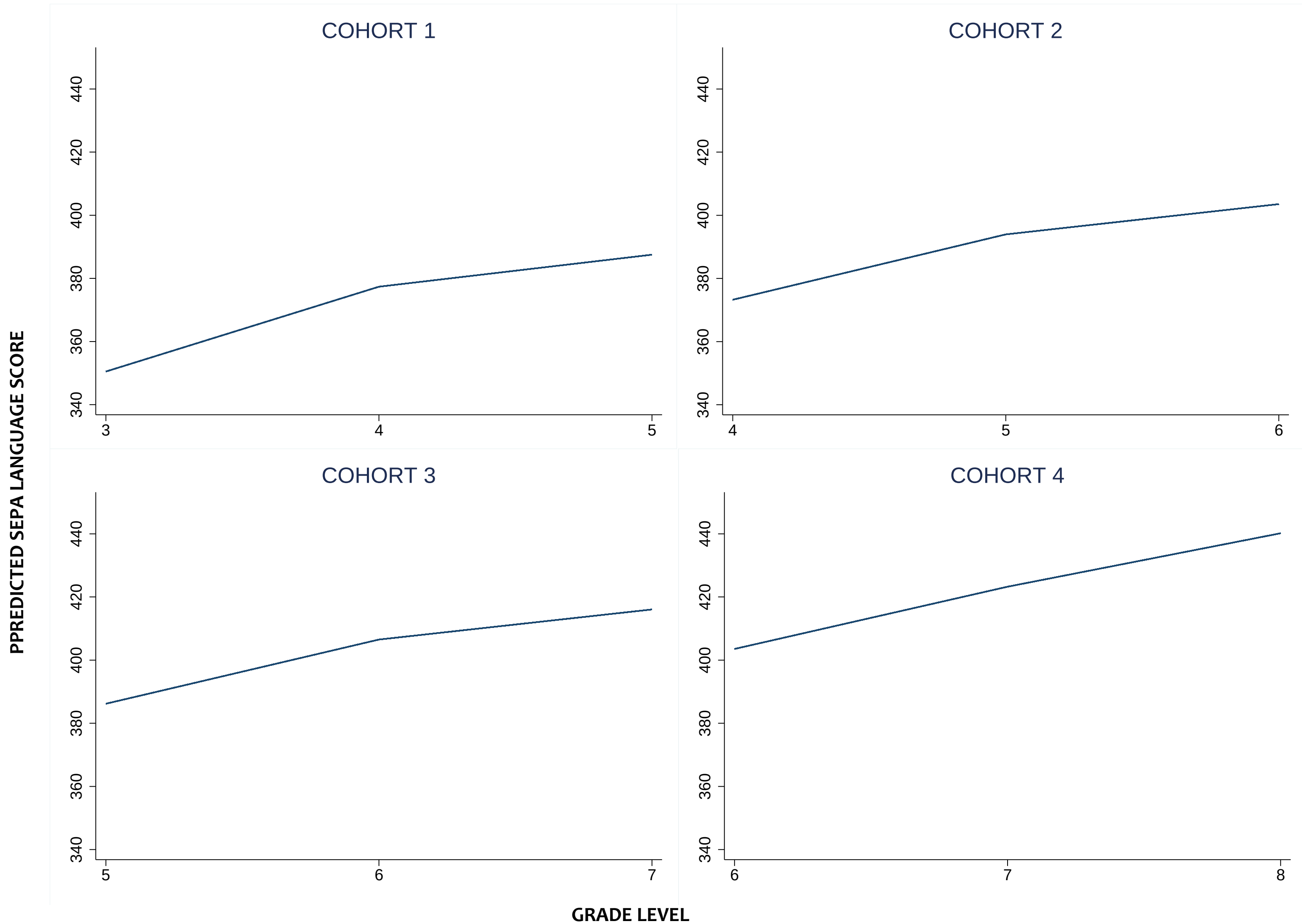
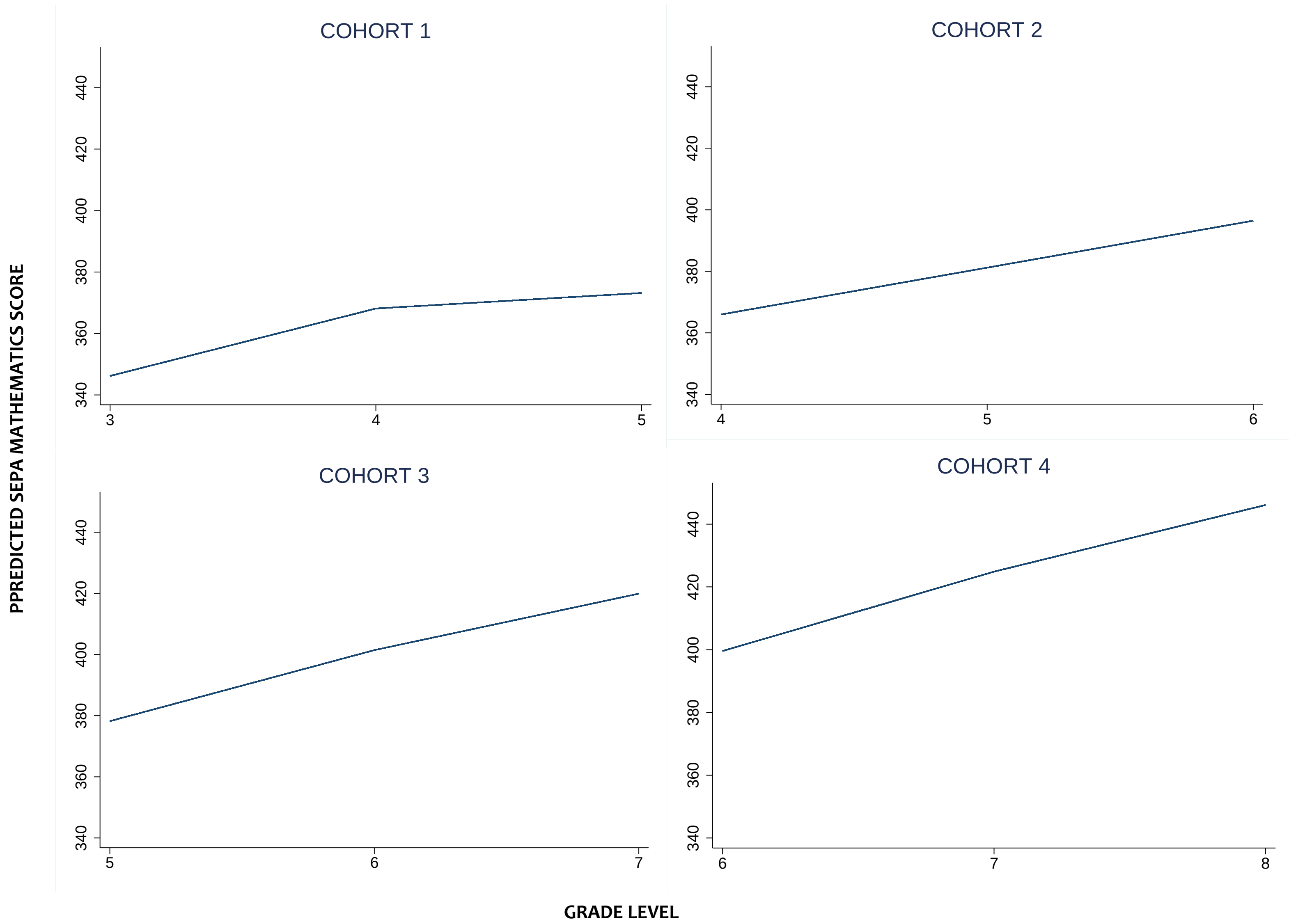


FIGURE 15: GROWTH CURVES FOR MATHEMATICS BASED ON MODEL 1, BY COHORT



The superiority of Model 1 over Model 0 is confirmed when the model fit indicators are compared. As summarised in Table 9, for all of the cohorts, with the only exception of Cohort 2 in mathematics, there are significant differences in the DIC indicators from Models 0 and 1, in favour of the latter. According to Spiegelhalter *et al* (2002), *DIC* differences greater than 7 are substantial, whereas significant differences here range from 123 to 3,280.

TABLE 9: DEVIANCE INFORMATION CRITERION (DIC) FOR MODELS 0 AND 1, BY COHORT

	DIC		
	MODEL 0		**MODEL 1**
LANGUAGE			
Cohort 1	141,925	>	138,771
Cohort 2	123,110	>	121,330
Cohort 3	103,135	>	101,860
Cohort 4	109,111	>	108,988
MATHEMATICS			
Cohort 1	141,230	>	137,950
Cohort 2	122,086	=	122,084
Cohort 3	100,864	>	100,596
Cohort 4	107,722	>	107,509

Note: Model 0 includes only a linear effect of time while Model 1 includes also a fixed quadratic effect of time.

In both Model 0 and Model 1 the results in terms of random effects suggest that there is significant variance in the intercept to be explained between students. Indeed, the variance of the intercept indicates that, for all of the cohorts, in both language and mathematics, there is a substantial difference in the mean score at the starting time point (grade level in year 2010) among students. In addition, the linear time slope varies across individuals in language and to a lesser extent in mathematics, indicating that growth rates at the starting point vary across students within cohorts.

The results above make it worthwhile to explore some characteristics of individuals to see how far they help to account for individual differences. In following sections, Model 1 is extended by adding further levels and student-level correlates.

4.4.2 *LINKAGE OF MULTIPLE COHORTS*

To test convergence of cohort-specific trajectories, the full model, that includes the separate mean trajectories for all cohorts simultaneously, and the reduced model, that includes a common trajectory for all cohorts over grades in the study, are estimated and compared. Table 10 shows the results of both models for language and mathematics.

Analysing the results for the full models it is possible to see that, as expected, all cohort effects for the level of achievement and for the slope are considerably different from zero, with the only exception of the slope term for Cohort 4 in mathematics. That is, all in all, Cohorts 2 to 4 tend to depart from Cohort 1 with respect to the level 1 intercept and the linear growth rate.

In addition, the estimated cohort effects for the level 1 intercept tend to increase for older cohorts (in language 22.819, 36.461, 53.237; in maths 19.896, 32.183, 53.540). These results confirm that cohort means in achievement increase with grade level. In language, the linear rate of increase within Cohort 1 is positive (34.118), and all of the other effects have negative sign (i.e., they present flatter trajectories, less accelerated growth than Cohort 1). However, even for Cohort 4, which has the largest negative cohort effect (-13.309), the estimated slope remains positive (i.e., 34.118 - 13.309 = 20.809). In mathematics, in turn, the linear rate of increase within Cohort 1 is also positive (30.154), but it decreases substantially in Cohort 2 and to a lesser extent in Cohorts 3 and 4.

As for the rate of acceleration, all of the cohorts exhibit rates different from that of Cohort 1. In language achievement, all of the cohorts, with the exception of Cohort 1 (-7.715), exhibit positive acceleration estimates, implying a departure from an arch-like curve to flatter, although still inverted-U shaped, curves. This indicates an increasing rate of change within cohorts. In mathematics, Cohort 1 also exhibits a negative acceleration estimate (-7.966). The change in the rate of acceleration is particularly pronounced in Cohort 2, where the curve becomes U-shaped, and then comes back to arch-like shapes in Cohorts 4 and 5 that are less pronounced than the one found in Cohort 1. This might be related to an important transition point located between Grade 4 (final year of first cycle of primary school) and Grade 5 (first year of the second cycle of primary school). As a consequence of this transition, students usually experience, and have to adjust to, a series of new school arrangements, such as, moving to a new building within the school and changing from being taught by the same teacher in all subjects to being taught by subject-specialist teachers.

In conclusion, the results show that students' growth in both language and mathematics is steep in initial periods but tends to decelerate over time. In the upper grades, academic growth in language is close to linear, and growth in mathematics achievement accelerates at the mid-point in the time series. Average growth rates for both language and mathematics are lower in the upper grades than in the lower grades. The fitted trajectories for the four cohorts are shown in Figure 16. The lines display the results in Table 10 and represent the four mean trajectories for each subject, based on the full model.

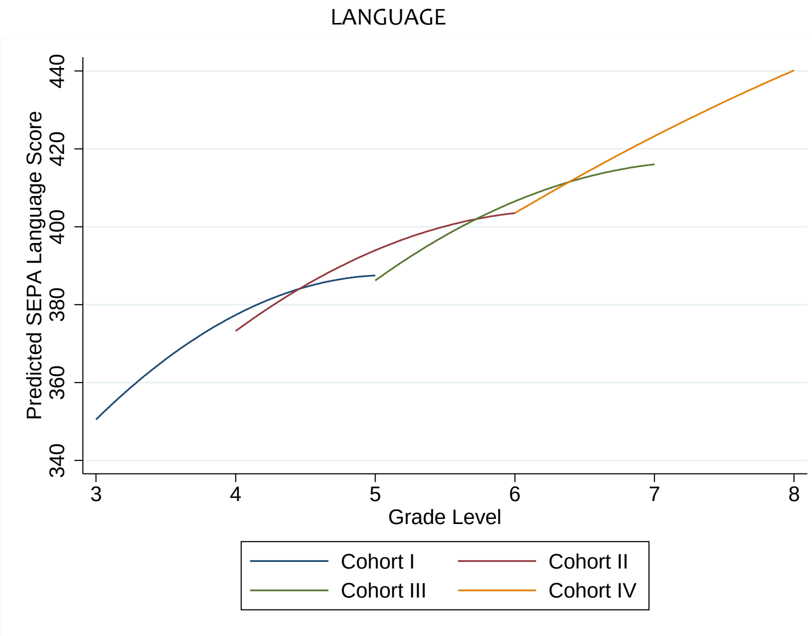
The fixed effects for the reduced model in language indicate that the polynomial function of time of degree 2 has a negative coefficient, which implies that the curve has a broad convex downward shape to it (indicating a deceleration in the outcome as a function of grade). The opposite is true for mathematics where there is a positive coefficient for the quadratic term, suggesting an overall acceleration in achievement as students progress through grade levels. On the basis of the results in Table 10, the mean trajectories obtained from the reduced model are graphed in Figure 17.

TABLE 10: FULL AND REDUCED MODEL 1 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 1 LANGUAGE				MODEL 1 MATHEMATICS			
	FULL MODEL		**REDUCED MODEL**		FULL MODEL		**REDUCED MODEL**	
	M (SE)	95% CI	**M (SE)**	**95% CI**	M (SE)	95% CI	**M (SE)**	**95% CI**
Fixed Part								
Intercept	350.602 (0.341)	349.914 - 351.290	**353.529 (0.225)**	**353.089 - 353.970**	346.169 (0.437)	345.211 - 347.126	**350.872 (0.672)**	**349.134 - 352.610**
Time	34.118 (0.949)	31.747 - 36.489			30.154 (1.778)	25.375 - 34.933		
$Time^2$	-7.715 (0.580)	-9.220 - -6.209			-7.966 (1.09)	-10.932 - -5.000		
Cohort 2	22.819 (0.439)	21.956 - 23.682			19.896 (0.475)	18.944 - 20.848		
Cohort 3	36.461 (0.689)	34.947 - 37.975			32.183 (0.514)	31.147 - 33.220		
Cohort 4	53.237 (0.480)	52.284 - 54.190			53.54 (0.595)	52.277 - 54.803		
Time x Cohort 2	-9.530 (0.783)	-11.207 - -7.853			-15.987 (1.753)	-20.530 - -11.443		
Time x Cohort 3	-9.908 (0.961)	-12.070 - -7.745			-5.32 (1.966)	-10.457 - -0.184		
Time x Cohort 4	-13.309 (0.899)	-15.302 - -11.315			-3.642 (2.136)	-9.283 - 1.998		
$Time^2$ x Cohort 2	3.031 (0.375)	2.228 - 3.835			8.618 (0.995)	5.991 - 11.245		
$Time^2$ x Cohort 3	2.867 (0.517)	1.660 - 4.073			5.956 (1.147)	2.903 - 9.010		
$Time^2$ x Cohort 4	6.418 (0.518)	5.205 - 7.631			6.401 (1.219)	3.135 - 9.666		
Time Cohort			**18.792 (0.261)**	**18.186 - 19.399**			**11.934 (0.656)**	**10.168 - 13.700**
Time $Cohort^2$			**-0.426 (0.047)**	**-0.534 - -0.317**			**1.358 (0.098)**	**1.099 - 1.617**
Random Part								
Random Level 2 (Student)								
Intercept	349.569 (8.692)	329.440 - 369.698	**338.844 (9.849)**	**315.321 - 362.367**	328.866 (7.799)	311.055 - 346.677	**324.840 (7.083)**	**309.098 - 340.582**
Time/Intercept	-28.799 (4.740)	-40.99 - -16.607	**-22.813 (5.197)**	**-36.269 - -9.357**	-7.512 (2.222)	-12.574 - -2.450	**-5.607 (1.665)**	**-9.187 - -2.028**
Time	11.462 (1.706)	7.520 - 15.404	**6.571 (2.021)**	**1.806 - 11.335**	2.283 (1.43)	-1.191 - 5.757	**0.777 (0.209)**	**0.364 - 1.189**
Random Level 1 (Occasion)								
Intercept	118.796 (2.044)	114.179 - 123.413	**135.531 (4.789)**	**123.013 - 148.049**	119.416 (1.766)	115.457 - 123.374	**132.300 (2.316)**	**349.134 - 352.610**
Units: Students	19,704		**19,704**		19,704		**19,704**	
Units: Occasions	59,112		**59,112**		59,112		**59,112**	
DIC	470,967		**478,081**		468,280		**474,641**	
pD	20,468		**18,048**		18,845		**17,494**	

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions and student levels. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

FIGURE 16: GROWTH CURVES BASED ON THE FULL MODEL 1



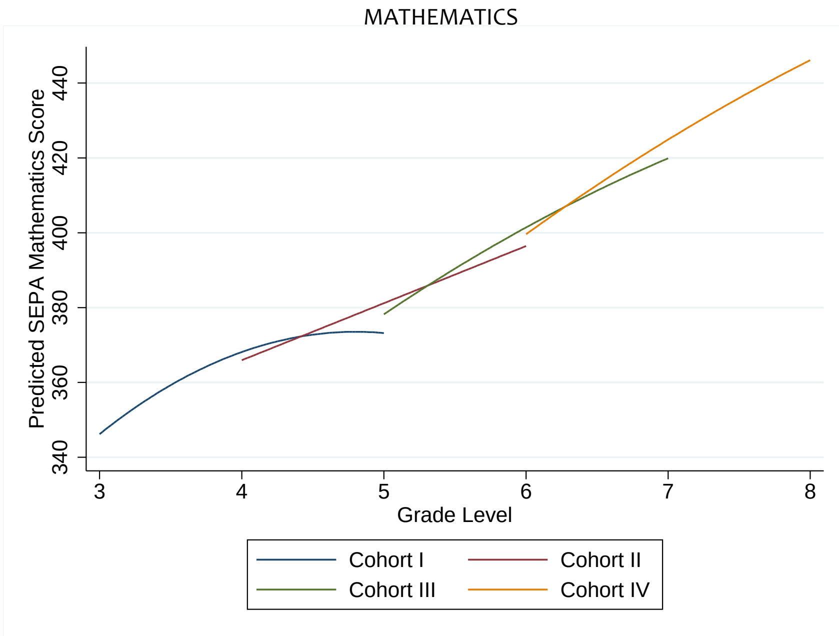
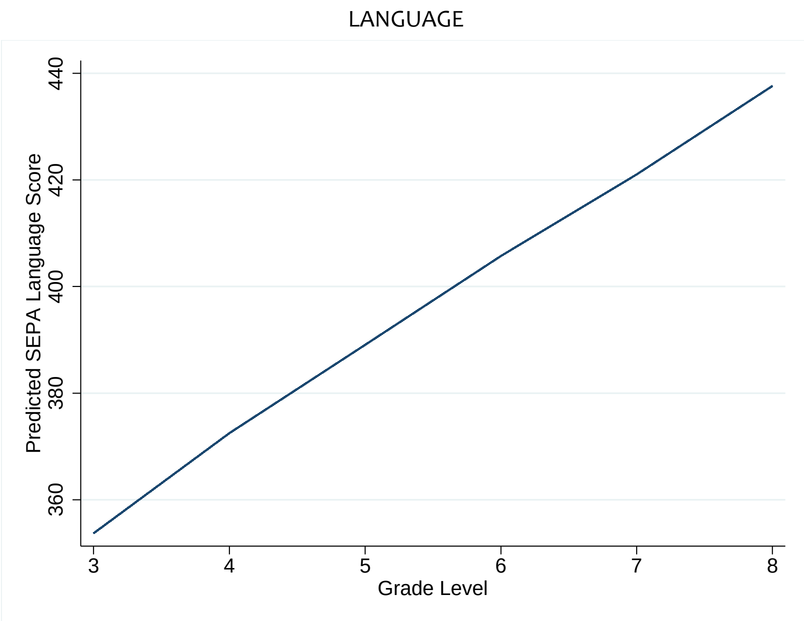
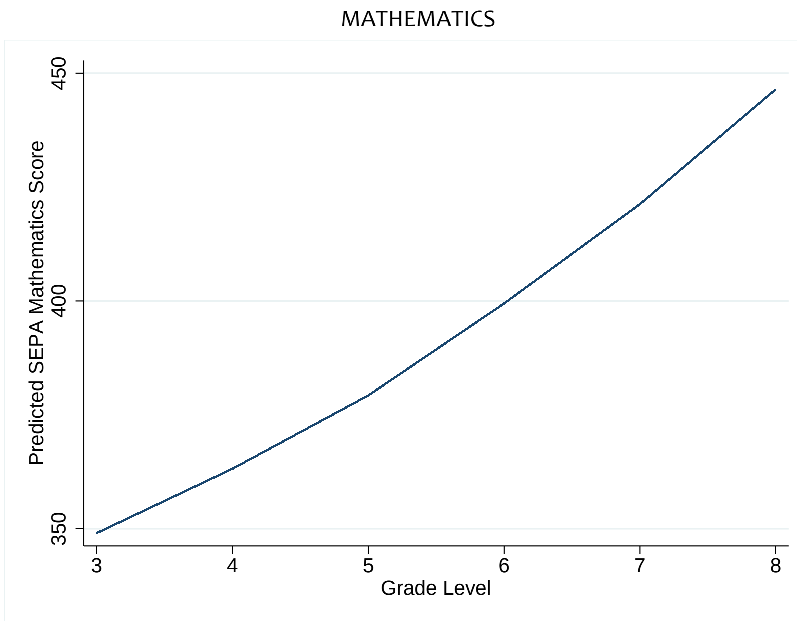


FIGURE 17: OVERALL GROWTH TRAJECTORY BASED ON THE REDUCED MODEL 1





Regarding the random effects, they are very similar across both specifications. The variance of the level 1 intercept is significantly greater than zero. Thus, there is significant variation among students within cohorts that is due to unmeasured person and perhaps school-specific characteristics. The existence of such variation encourages interest in the possible effects of covariates related to individual differences in trajectories of change.

After estimating the models, the convergence of cohort-specific trajectories is tested by comparing the deviance of the full and the reduced models. In language, the deviance associated with the reduced model is $DIC_{reduced}$ = 478,081. This deviance is larger than the deviance of the full model, DIC_{full} = 470,967, with a difference of ΔDIC = -7,114 in favour of the full model. In mathematics, the deviance associated with the reduced model is $DIC_{reduced}$ = 474,641. As in language, this deviance is larger than the deviance of full model is DIC_{full} = 468,280, this time with a difference of ΔDIC = -6,361.

The magnitudes of the reduction in deviance associated with the full model are large enough to suggest that the additional cohort parameters are helpful in accounting for the data. Therefore, based on the Deviance Information Criterion (DIC) for MCMC estimation methods, it was decided to reject the null hypothesis that the cohort effects are ignorable. As cohort effects are found to be significant, it was decided that these effects must be included in future models. Substantively, this means that there are some systematic sub-sample differences between the cohorts. Thus, all inferences regarding correlates of change are adjusted for cohort membership.

4.4.3 *Decomposing Variance at the School and Student Levels*

In the previous section the quadratic nature of language and mathematics achievement development from the end of Grade 3 to the end of Grade 8 in primary school was depicted. Differences in the growth trajectories for these two outcomes were also identified. These results are in line with the literature, which shows an overall decrease in the rate of growth in achievement as students move to the later years of primary school. Analyses also showed that individuals differ in their achievement status and, to some extent, in their rate of development over time.

In this section, the hierarchical model 2, which introduces school-level variation on student achievement status and growth (Equation 3), is analysed in order to differentiate this source of variance from that arising from the student level. This multilevel model consists of three levels: schools, students within schools, and time points within students. By means of this empty growth model (no other covariates added, explanatory variables will be added in later models), the raw unadjusted differences between schools (school variance) can be estimated. Cohort effects are added, as specified in the full model. The main findings regarding school effects are presented in Chapter 5.

In Figure 18, the mean development of language and mathematics achievement for each of the 156 schools in the final sample is depicted. These graphs provide a first approach to the shape of growth at the school level. Schools seem to vary both in terms of their intercepts and slopes and, all in all, they seem to have curvilinear growth trajectories.

As shown in Table 11, the addition of school random effects significantly improves the model fit as the DIC decreases from 470,967 to 467,989 in language (ΔDIC = -2,978), and from 468,280 to 463,418 in mathematics (ΔDIC = -4,862). This shows that growth trajectories differ significantly between schools.

FIGURE 18: SCHOOL MEAN ACHIEVEMENT TRAJECTORIES IN LANGUAGE AND MATHEMATICS

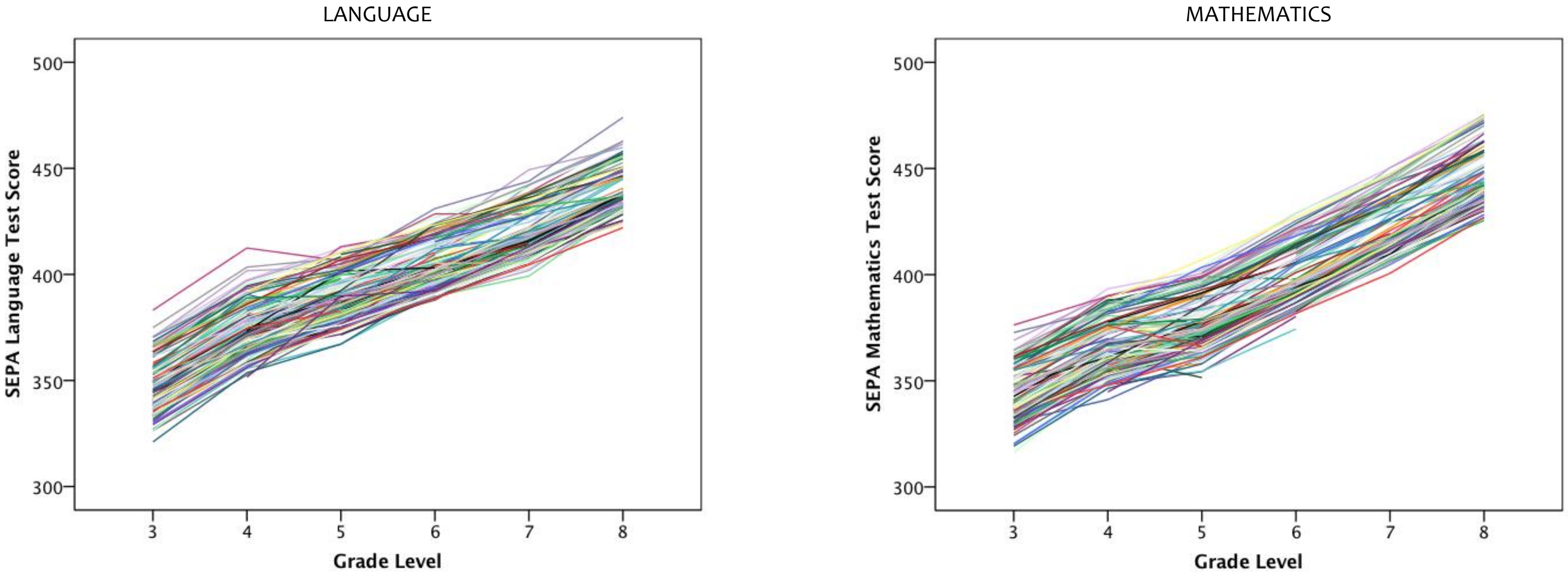


TABLE 11: FULL MODEL 2 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 2 LANGUAGE		MODEL 2 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Fixed Part				
Intercept	348.995 (1.046)	346.942 - 351.049	343.627 (1.128)	341.404 - 345.849
Time	34.665 (1.155)	31.783 - 37.547	30.578 (1.974)	25.296 - 35.861
$Time^2$	-7.714 (0.580)	-9.219 - -6.208	-7.966 (1.089)	-10.933 - -4.999
Cohort 2	25.238 (0.405)	24.442 - 26.033	23.179 (0.487)	22.161 - 24.197
Cohort 3	36.627 (0.514)	35.579 - 37.676	33.423 (0.465)	32.479 - 34.366
Cohort 4	54.510 (0.459)	53.600 - 55.421	55.941 (0.508)	54.885 - 56.998
Time x Cohort 2	-9.950 (0.796)	-11.661 - -8.238	-16.655 (1.959)	-21.805 - -11.504
Time x Cohort 3	-9.951 (1.060)	-12.403 - -7.498	-5.920 (2.177)	-11.675 - -0.164
Time x Cohort 4	-13.874 (1.021)	-16.222 - -11.525	-4.529 (2.358)	-10.817 - 1.759
$Time^2$ x Cohort 2	3.030 (0.375)	2.226 - 3.834	8.617 (0.993)	5.989 - 11.246
$Time^2$ x Cohort 3	2.866 (0.517)	1.660 - 4.073	5.956 (1.146)	2.901 - 9.011
$Time^2$ x Cohort 4	6.417 (0.518)	5.204 - 7.630	6.400 (1.218)	3.133 - 9.667
Random Part				
Random Level 2 (School)				
Intercept	146.598 (19.751)	107.251 - 185.944	160.403 (19.161)	122.824 - 197.983
Time/Intercept	-21.078 (5.412)	-32.650 - -9.506	-12.742 (3.936)	-20.589 - -4.895
Time	9.836 (1.933)	5.609 - 14.064	9.154 (1.524)	5.992 - 12.316
Random Level 2 (Student)				
Intercept	209.528 (3.575)	202.446 - 216.611	172.101 (2.970)	166.209 - 177.994
Time/Intercept	-9.314 (1.587)	-12.620 - -6.008	-2.838 (0.951)	-4.751 - -0.925
Time	3.115 (0.877)	1.370 - 4.861	0.562 (0.150)	0.267 - 0.857
Random Level 1 (Occasion)				
Intercept	118.934 (2.017)	114.334 - 123.534	113.886 (1.218)	111.268 - 116.503
Student Initial Status – Growth Correlation	-0.365		-0.289	
School ICC Initial Status	0.412		0.482	
School ICC Growth	0.759		0.942	
Units: Schools	156		156	
Units: Students	19,704		19,704	
Units: Occasions	59,112		59,112	
DIC	467,989		463,418	
pD	17,359		16,494	

Model 2 predicts an average language and mathematics proficiency of 348.995 and 343.627, respectively, in Grade 3 (see Table 11). Students grow significantly in their language and mathematics skills as they advance in school from Grade 3 to 8. Measured by the growth rate coefficient, in Cohort 1, they grow in 34.665 and 30.578 on average every year in language and mathematics, respectively. In language, the growth rates for Cohorts 2, 3 and 4 reduce importantly when compared to that in Cohort 1 (growth rates of 24.715, 24.714, 20.791 for Cohorts 2, 3 and 4, respectively). In mathematics, growth rates decrease substantially in average in Cohort 2 (13.923) and then increase steadily in cohorts 3 and 4 (growth rates of 024.658 and 26.049 for Cohort 3 and 4, respectively). Scores within cohorts do not grow at a constant rate of change over this period, though. The negative estimates of the acceleration parameter observed for both subjects in every cohort, with the exception of Cohort 2 in mathematics, indicate a curvilinear growth trajectory, this is, that students grow in their language and math skills at a decreasing rate of change.

Model estimates confirm that variation around the grand mean of the initial status and, to a lesser extent, of the growth rate for students are still substantial. This is, after controlling for school-level variation, there is significant variance between children at the end of third grade and in their growth rates. The student-level standard deviation at entry[29] is 14.475 and 13.119, for language and mathematics, respectively, indicating wide variability in students' incoming attainment levels. The student-level standard deviation for the academic year learning rate[30] is 1.765 and 0.750, for language and mathematics, respectively indicating some variability in learning among children after controlling for their specific schools.

Finally, the random effects are moderately and negatively correlated (-0.365 in language and -0.289 in mathematics)[31], indicating that children who start with lower proficiency levels tend to learn at a faster rate than those who commence with higher proficiency

[29] Calculated as $\sqrt{\boldsymbol{\sigma}^2_{\boldsymbol{u0}}}$.

[30] Calculated as $\sqrt{\boldsymbol{\sigma}^2_{\boldsymbol{u1}}}$.

[31] The correlation can be derived from the estimates of the variance and covariance components. Thus, in the case of this model: $r_{u0u1} = \frac{\sigma_{u0u1}}{\sqrt{(\sigma^2_{u0} \times \sigma^2_{u1}}}$.

levels. Thus, the achievement gap seems to narrow down as students progress through primary education. This suggests a general fanning in pattern of student trajectories (i.e., the starting points are more spread out than the end points). The finding is important as it suggests that initial low attainers begin to catch up over time in both subjects.

A negative relationship between initial status and growth might be due to a source of bias that needs to be considered when estimating individual measures of achievement growth, namely, ceiling effects in achievement growth. These effects are present when the tests are not sufficiently difficult to capture growth of the high-achieving students and, therefore, they might prevent us from measuring achievement gaps and educational effects (Koedel and Betts, 2008; Martineau, 2006). Potential ceiling effects were examined and, as shown in Appendix 5, it was found that the proportion of students at each grade level that achieve the maximum test score is negligibly small ranging from 0.031% to 0.169% in language and from 0.033% to 0.641% in mathematics. Also, scores on each of the three measurement points, for the four cohorts and in both subjects are approximately normally distributed (see distribution indicators in Appendix 5). Within the private school sector, the one with the highest SES levels, the language and mathematics achievement distributions also fit nicely a normal distribution, and less than 3% of the students achieved the maximum score in these measures. Hence, the SEPA achievement tests distinguish well among students in their language and mathematics achievement levels and there does not seem to be an indication of ceiling effects operating in the data.

Model 2 was also fitted separately for each of the different types of schools (i.e., public, private subsidised and private non-subsidised schools) in order to determine whether the negative relationship between initial status and growth is characteristic of some of these school sectors only. The results, presented in Appendix 6, show that, although the negative correlations between initial status and growth are of higher magnitude in the private sector, they are generally present across school sectors (in language, -0.136 for public schools, -0.502 for private subsidised schools and -0.562 for private non-subsidised schools; in mathematics, 0.036 for public schools, -0.600 for private subsidised schools and -0.020 for private non-subsidised schools).

4.4.4 *Adding Student-Level Predictors*

The results presented so far indicate significant variation among students, both, in terms of intercepts and rates of achievement, encouraging a search for differences in students' background and demographic characteristics that might account for these variations. Thus, Model 3 was fitted using student gender, age[32], family socio-economic status (SES) and number of books at home as predictors.

Results are shown in Table 12 and indicate that both initial score and rate of change are predicted by student characteristics. Indeed, there is a sizeable improvement in the model fit when student-level covariates are added, as shown by the reduction on DIC values from Model 2 to Model 3 (ΔDIC = -72.795 in language, and ΔDIC = -161.400 in mathematics). Nonetheless, these student variables are able to explain only 3.54% and 3.48% of the student-level variance in language and mathematics achievement status, respectively, and an even lower percentage of the variance on student achievement growth.

Interestingly, these variables explain an important proportion of the school-level variance in status at time 0 (27.55% in language and 25.20% in mathematics). The fact that student covariates explain a large proportion of the school variance in achievement is a clear indication of marked differences in school intakes, in terms of these variables. Indeed, as other authors have previously noticed, the proportion of SES variation that occurs between schools is extremely large in Chile, reaching 62% (Mizala and Torche, 2012), indicating that the school is a pivotal unit of stratification.

As shown in Table 12, all of the student-level predictors are related to the intercept in both subjects. Being female is positively related to achievement status in language and negatively associated with achievement status in mathematics at time 0. The differences are small, however. Considering for example Grade 3, the standard deviation of language and mathematics scores is 21 and 20, respectively, while the mean differences between boys and girls, after controlling for the rest of the student variables, are only of two points in both subjects. It is important to note that girls not only have lower achievement

[32] Age was calculated in years and months and grand-mean centered within cohorts.

levels than boys in math, but they also grow more slowly in their math skills, indicating an increase in the gender gap over time. This trend is depicted in Figure 19.

Being older within each cohort is negatively related to the intercept. When the rest of the covariates at the student level are controlled for, being one year older reduces by 2 points the achievement scores in both subjects. In language, older students within their cohorts experience a significantly lower growth rate, this is to say, there is a flatter slope in language achievement for older students. These age effects are likely to reflect differences in achievement status and growth rates for those students in the sample who experienced grade retention prior to 2010, the first wave considered in this study, and became the oldest students within their new year groups[33].

Family SES is positively related to achievement levels. For an increment of 1 unit in the SES index, language and math achievement increase in 0.772 and 1.168, respectively. However, results of Model 3 also indicate that family SES does not contribute significantly to growth rates in either of the subjects. These results suggest that high SES students start with higher achievement that their less advantaged peers and that gaps between students from varying socioeconomic backgrounds are not reduced nor increased as students progress through primary school. In Figure 21 this trend is depicted, showing that relative positions of students in different SES quartiles remain stable across time.

Finally, irrespective of family SES, children whose parents report a higher number of books at home, perform better in both language and mathematics. Differences of 1 unit in this variable, when the rest of the covariates at the student level are controlled, amount to an increase of 2.460 and 2.078 on language and mathematics achievement, respectively. However, there are no apparent differences in growth rates related to the reported number of books at home.

[33] Previous research has shown that older students within a year group do better than younger ones, unless students are retained due to low attainment. However, in this study, the effect of grade retention is not distinguishable from that of age.

TABLE 12: FULL MODEL 3 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 3 LANGUAGE		MODEL 3 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Fixed Part				
Intercept	347.750 (0.919)	345.945 - 349.554	344.462 (1.016)	342.453 - 346.471
Time	34.582 (1.089)	31.899 - 37.266	31.056 (1.952)	25.842 - 36.271
Time2	-7.728 (0.574)	-9.214 - -6.241	-7.995 (1.090)	-10.965 - -5.026
Cohort 2	25.138 (0.398)	24.357 - 25.919	23.142 (0.481)	22.137 - 24.147
Cohort 3	36.904 (0.507)	35.871 - 37.938	33.702 (0.442)	32.817 - 34.586
Cohort 4	54.790 (0.447)	53.905 - 55.675	56.241 (0.476)	55.267 - 57.215
Time x Cohort 2	-10.058 (0.797)	-11.772 - -8.344	-16.780 (1.961)	-21.938 - -11.623
Time x Cohort 3	-10.032 (1.085)	-12.559 - -7.506	-6.037 (2.175)	-11.791 - -0.283
Time x Cohort 4	-13.966 (1.009)	-16.277 - -11.655	-4.649 (2.363)	-10.952 - 1.653
Time2 x Cohort 2	3.083 (0.376)	2.278 - 3.888	8.685 (0.994)	6.055 - 11.316
Time2 x Cohort 3	2.881 (0.523)	1.658 - 4.104	5.983 (1.145)	2.927 - 9.038
Time2 x Cohort 4	6.437 (0.513)	5.237 - 7.638	6.433 (1.220)	3.160 - 9.705
Female student	2.095 (0.309)	1.475 - 2.715	-2.008 (0.311)	-2.645 - -1.371
Age	-1.737 (0.271)	-2.274 - -1.200	-1.736 (0.281)	-2.308 - -1.164
SES	0.772 (0.308)	0.050 - 1.495	1.168 (0.307)	0.437 - 1.899
Number books at home	2.460 (0.152)	2.162 - 2.758	2.078 (0.165)	1.743 - 2.413
Time x Female	0.359 (0.202)	-0.096 - 0.814	-0.695 (0.151)	-1.010 - -0.381
Time x Age	-0.334 (0.133)	-0.603 - -0.065	-0.463 (0.207)	-0.949 - 0.024
Time x SES	-0.057 (0.281)	-0.760 - 0.645	-0.268 (0.135)	-0.547 - 0.011
Time x Number books at home	-0.223 (0.110)	-0.468 - 0.022	-0.088 (0.083)	-0.261 - 0.084
Random Part				
Random Level 3 (School)				
Intercept	106.206 (14.932)	76.328 - 136.083	119.987 (15.331)	89.784 - 150.189
Time/Intercept	-17.489 (4.070)	-25.810 - -9.169	-10.555 (3.486)	-17.504 - -3.605
Time	9.757 (1.807)	5.895 - 13.618	9.247 (1.523)	6.096 - 12.398
Random Level 2 (Student)				
Intercept	202.104 (3.490)	195.218 - 208.991	166.116 (3.102)	159.852 - 172.381
Time/Intercept	-9.184 (1.623)	-12.54 - -5.828	-3.195 (0.935)	-5.087 - -1.302
Time	2.927 (0.930)	1.092 - 4.763	0.586 (0.165)	0.263 - 0.910
Random Level 1 (Occasion)				
Intercept	119.151 (2.067)	114.473 - 123.828	113.744 (1.237)	111.074 - 116.414

	MODEL 3 LANGUAGE		MODEL 3 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
School Initial Status – Growth Correlation	-0.543		-0.317	
Student Initial Status – Growth Correlation	-0.378		-0.324	
School ICC Initial Status	0.344		0.419	
School ICC Growth	0.769		0.940	
DIC	467,900		463,219	
pD	17,220		16,382	
Units: Schools	156		156	
Units: Students	19,704		19,704	
Units: Occasions	59,112		59,112	

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions, student and school levels. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

FIGURE 19: GROWTH CURVES BY STUDENT GENDER BASED ON MODEL 3

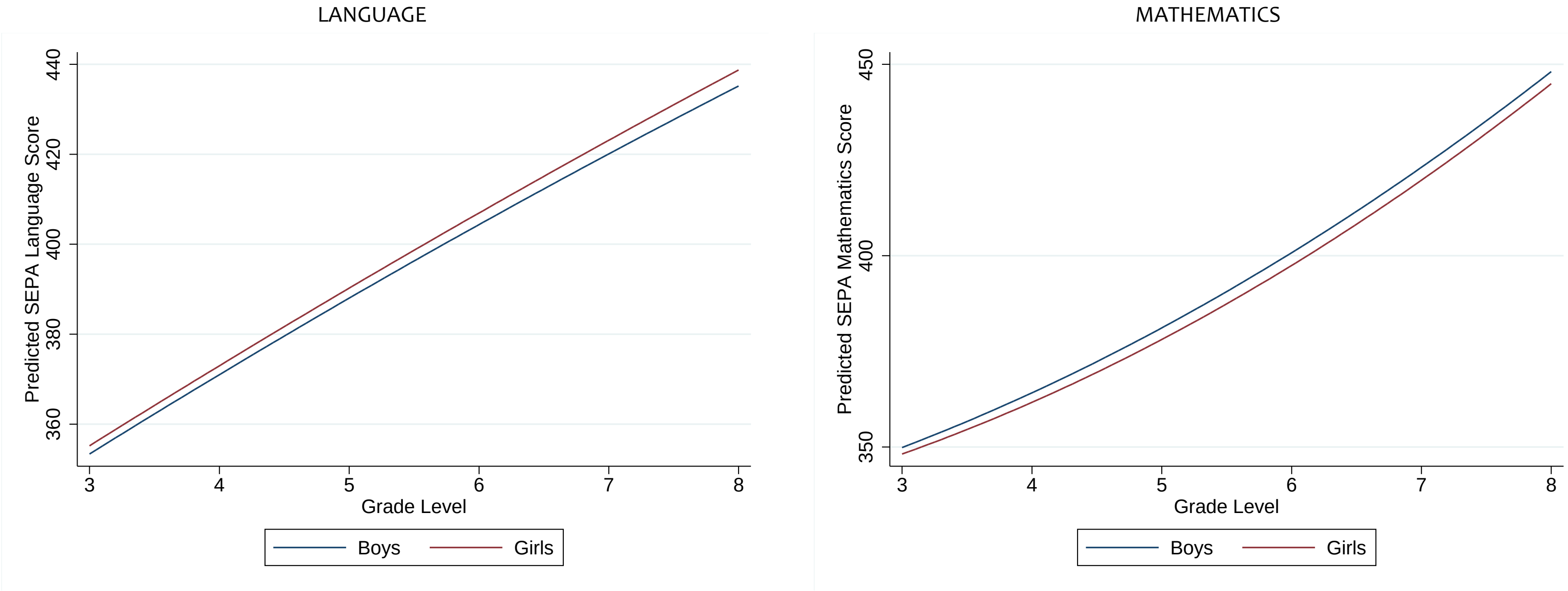


FIGURE 20: GROWTH CURVES BY STUDENT WITHIN-COHORT AGE QUARTILES BASED ON MODEL 3

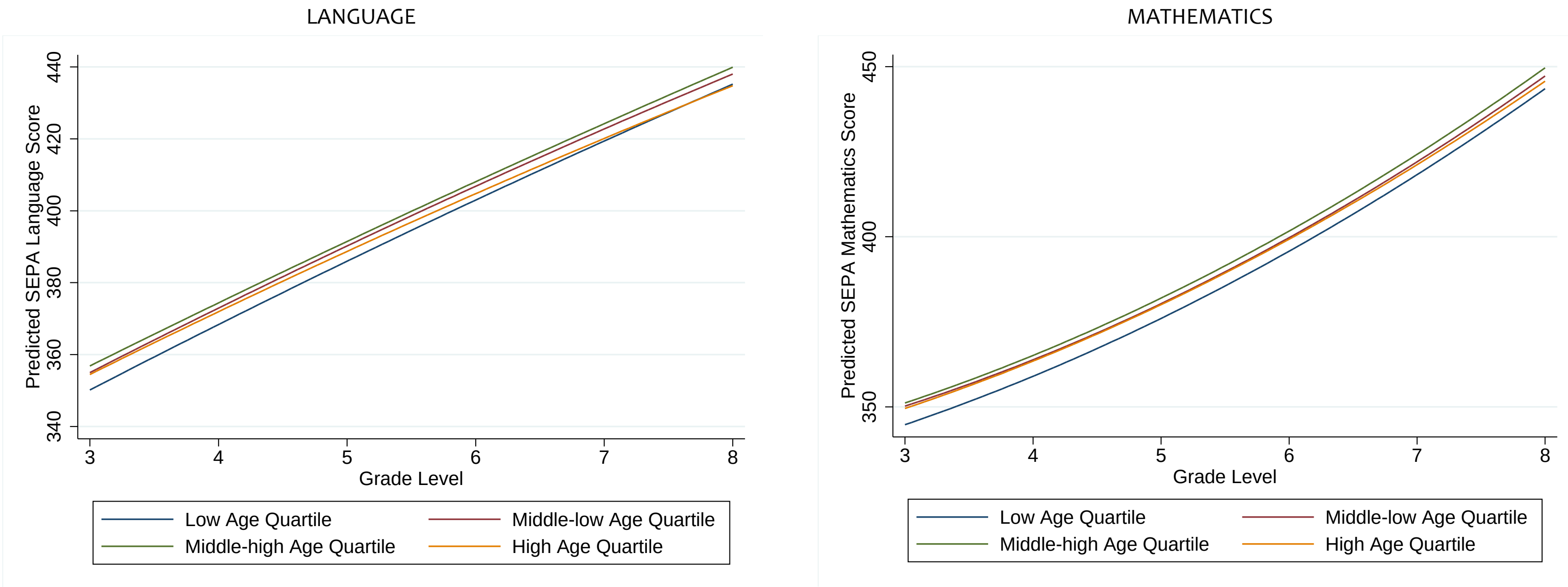


FIGURE 21: GROWTH CURVES BY FAMILY SES QUARTILES BASED ON MODEL 3

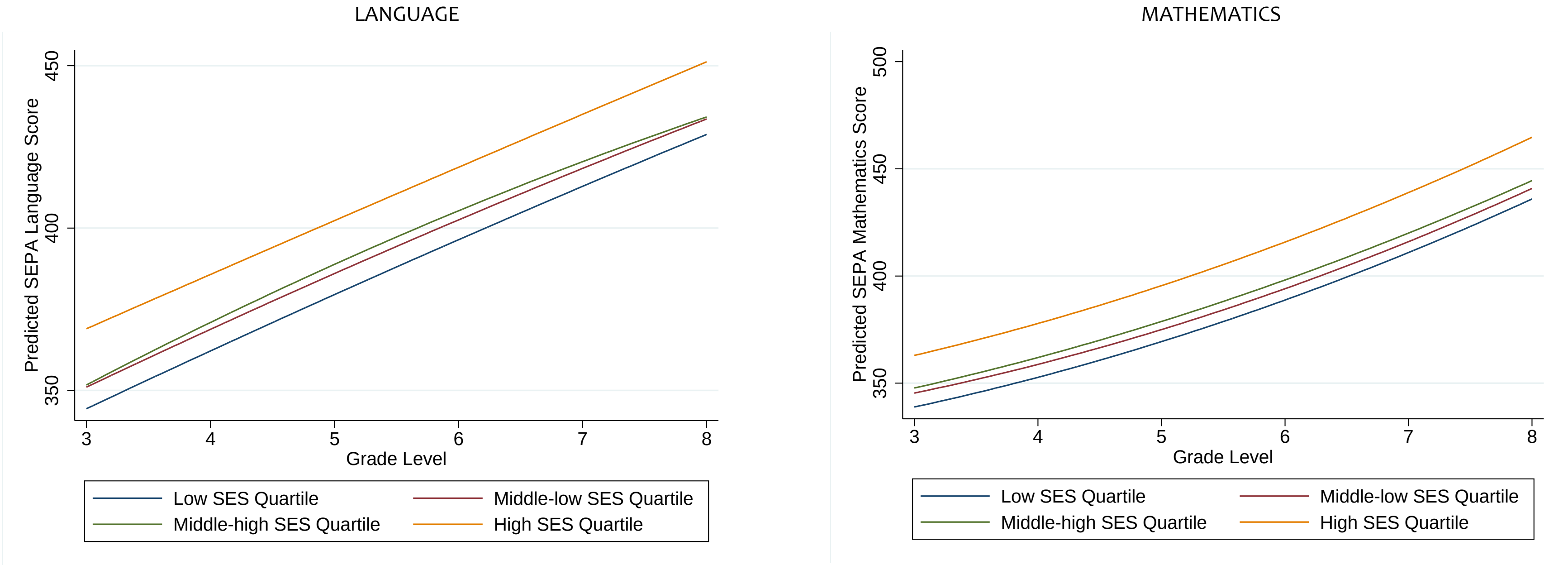
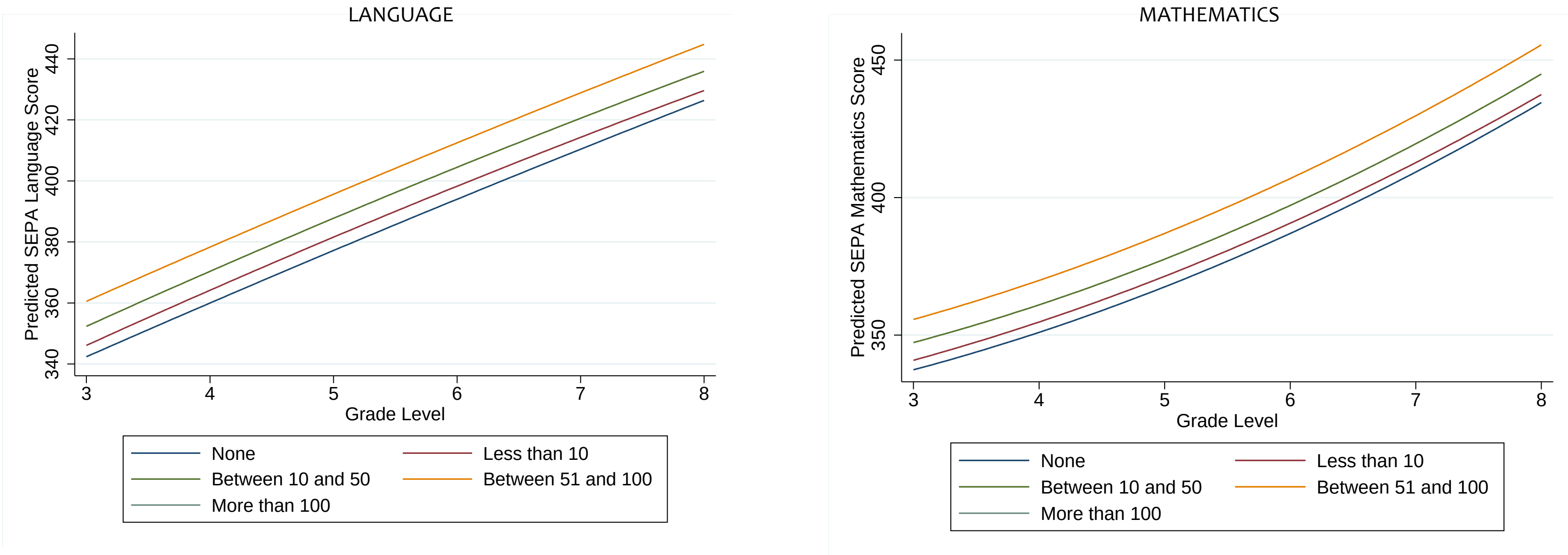


FIGURE 22: GROWTH CURVES BY RANGES OF NUMBER OF BOOKS AT HOME BASED ON MODEL 3



4.5 Conclusions

In this chapter the convergence of expected trajectories for students' language and mathematics achievement into an accelerated longitudinal design was tested. In the study's data, cohort effects might arise because of the varied demographic composition of the cohorts resulting from demographic and historical changes in the country, but also because of changes in the sample of schools participating in the SEPA project across the three years of interest. Thus, the effects of history and selection in this study may be confounded. Cohort effects were found to be substantial and therefore have been retained in subsequent models.

Also, the multilevel approach applied to a three time point longitudinal design enabled to separate student and school variation from variation due to test-level measurement error while drawing on a substantial source of intra-individual variation to examine change. Thus, this methodological examination is innovative and advances prior research.

Turning back to the research questions addressed in this chapter, the results can be summarised as follows. The study shows that both students' language and math proficiency show upward growth. Furthermore, growth of student achievement does not appear to be linear but a curvilinear process. Analyses also showed that individuals differ both in their achievement status and rate of development over time.

The use of three time points and statistical techniques well suited to the accelerated longitudinal data also contributed to the methodological advance of measurement of the academic achievement gap trends. Each student-level characteristic is significantly related to the outcome in the expected direction based on the literature. Boys perform slightly better in math and worse in language, signalling a "gender division of learning" similar to most countries in the world (Ma, 2008). In math, boys present higher growth rates than girls, making the gender gap grow from 3rd to 8th grade.

Surprisingly, the findings indicate that the achievement gap between high- and low-socioeconomic status students remains fairly stable over the course of primary school years. Some previous investigations on the achievement gap related to family SES that have used sound longitudinal designs and advanced statistical methods have arrived to similar results (e.g., Caro, McDonald and Willms, 2009; Wilkins and Ma, 2002). Other variables that were found to predict student achievement status were number of books at home and within-cohort age. With the exception of gender, the patterns of effects are very similar for language and mathematics, indicating that the results are not an artifact of a particular subject.

All of the relevant features of students' achievement trajectories found in this chapter are carried forward to the models introduced in the following chapters, in which the magnitude, consistency, predictors and differentiation across student groups of school and teacher effects are investigated.

CHAPTER 5: SCHOOL EFFECTS ON STUDENT ACHIEVEMENT TRAJECTORIES

5.1 INTRODUCTION

In the previous chapter the non-linear shape of language and mathematics achievement development from the end of Grade 3 to the end of Grade 8 in primary school was depicted. Differences in the growth trajectories for these two outcomes were also reported. These results are in line with the literature, which shows an overall decrease in the rate of growth in achievement as students move to the later years of primary school. Analyses also showed that individuals differ in their achievement status and, to some extent, in their rate of development over time, and that the effects of students' demographics and background characteristics are stronger when attainment at any one time point is analysed, but they only explain a negligible proportion of the variance on student achievement growth rates. This section explores school effects on student achievement status and growth.

5.2 RESEARCH QUESTIONS

The second aim of this study is to investigate school effects on student achievement, by analysing their magnitude, compositional effects, consistency across subjects (i.e., language and mathematics) and differential impact across student groups. This aim unfolds into the following research questions, to be answered in this chapter:

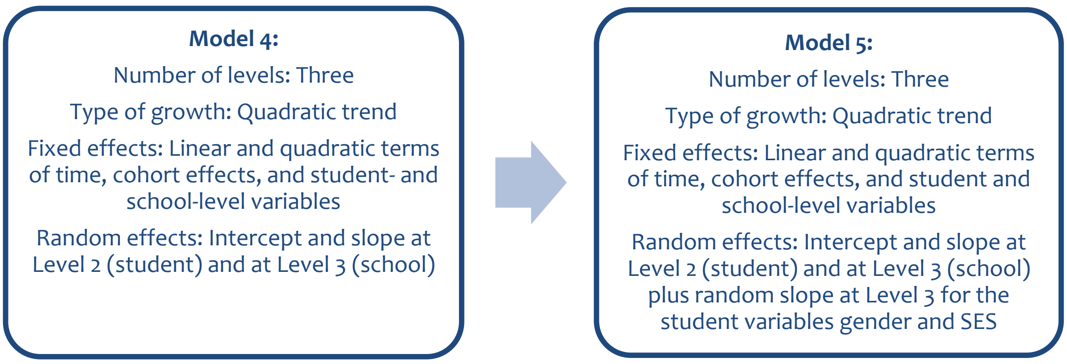
1. How large are the effects of schools on the achievement of their students and on the changes in achievement over time?
2. Are there school compositional effects on educational achievement status and growth?
3. How consistent are school effects across different academic subjects (i.e., language and mathematics)?
4. Are schools differentially effective across student groups?

5.3 METHOD OF ANALYSIS

5.3.1 *MODELS*

Figure 23 shows the progression of models that will be explored in this chapter. The analyses are organised in two models. Model 4 is a within-school model, with student and school characteristics. Model 5 adds random effects at the school level for the student variables gender and SES.

FIGURE 23: GROWTH MODELS IN CHAPTER 5



5.3.1.1 *MODEL 4. A THREE-LEVEL ACCELERATED GROWTH MODEL WITH A QUADRATIC EFFECT OF TIME, AND STUDENT- AND SCHOOL-LEVEL COVARIATES*

The estimation of generic school effects is based on Model 4, a between-school model with the same random-effects specification as Model 3 in Chapter 4, but with the addition of school-level characteristics. The importance of controlling for student background and school characteristics in statistical models, before comparisons across schools and teachers can be made, is well established in the field of educational effectiveness research (Aitkin and Longford, 1986; Sammons and Luyten, 2009; Willms, 2010). Model 4, which also includes the cohort effects components of the full model, is depicted in Equation 8.

The intraclass correlation (ICC) is considered a measure of overall size of school effects in EER. However, in a three-level random intercept and slopes model, such as Model 4, the definition of the ICC becomes trickier resulting in inconsistencies between different authors (Singer and Willett, 2003). In this study, the calculation of school effects in a linear growth curve model with a random intercept and slope made by Raudenbush and Bryk (2002) is used. The authors recommend using the percentage of variation that lies between schools for both the initial status and growth to measure the school effects.

EQUATION 8

$$\boldsymbol{Y_{tij} = \beta_{0ij} + \beta_{1ij}t_{tij} + \beta_2 t_{tij}^2 + e_{tij},}$$

$$\boldsymbol{\beta_{0ij} = \beta_{00j} + \beta_{010}Cohort2_{ij} + \beta_{020}Cohort3_{ij} + \beta_{030}Cohort4_{ij} + \beta_{040}Female_{ij} + \beta_{050}SES_{ij} + \beta_{060}Books_{ij} + \beta_{070}Age_{ij} + u_{0ij}}$$

$$\boldsymbol{\beta_{1ij} = \beta_{10j} + \beta_{110}Cohort2_{ij} + \beta_{120}Cohort3_{ij} + \beta_{130}Cohort4_{ij} + \beta_{140}Female_{ij} + \beta_{150}SES_{ij} + \beta_{160}Books_{ij} + \beta_{170}Age_{ij} + u_{1ij}}$$

$$\boldsymbol{\beta_{00j} = \beta_{000} + \beta_{200}Achievement\ Mean_j + \beta_{300}Achievement\ SD_j + \beta_{400}School\ SES_j + u_{00j}}$$

$$\boldsymbol{\beta_{10j} = \beta_{100} + \beta_{500}Achievement\ Mean_j + \beta_{600}Achievement\ SD_j + \beta_{700}School\ SES_j + u_{01j}}$$

$$\boldsymbol{\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]}$$

$$\boldsymbol{\begin{pmatrix} u_{00j} \\ u_{01j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u00}^2 & \\ \sigma_{u00u01} & \sigma_{u01}^2 \end{pmatrix}\right]}$$

$$\boldsymbol{e_{tij} \sim N(0, \sigma_e^2).}$$

Model 4 is also used to analyse the effect of school composition. As shown in Equation 8, school-level predictors are included in Model 4 to depict compositional effects, referring to the influence of the makeup of the school body on student achievement status and growth. The school-level variables used are *Achievement Mean,* indicating school mean score on the SIMCE Assessment System test for the relevant subject (i.e., reading or mathematics), *Achievement SD,* referring to the within-school standard deviation in SIMCE test scores for the relevant subject, a measure of diversity in the levels of achievement of the student body, and *School SES*, as calculated by the SIMCE

Assessment System[34]. It is important to note that these three variables are calculated using data from a different source from that of the dependant variables (the SEPA Assessment Project). This represents an advantage of the present study as compositional effects are often measured by aggregating individual-level variables to form school-level variables, which is likely to lead to biased estimations (Televantou et al., 2015). Appendix 1 provides a description of the SIMCE Assessment System and Appendix 4 shows the operational definition and descriptive statistics for the school-level variables. The student variable *SES* at level 2 was grand-mean centered such that the school SES coefficient captures the composition effect directly (Enders and Tofighi, 2007; Raudenbush and Bryk, 2002).

Model 4 also elicits the analysis of consistency of school effects by studying the correlation between school effects on language and mathematics. In order to assess whether schools have consistent effects on students' achievement trajectories across academic subjects, correlation coefficients between school intercept and slope residuals in language and mathematics were calculated for the schools in the sample.

5.3.1.2 *MODEL 5. A THREE-LEVEL ACCELERATED GROWTH MODEL WITH A QUADRATIC EFFECT OF TIME, STUDENT- AND SCHOOL-LEVEL COVARIATES AND A RANDOM EFFECT OF STUDENT CHARATERISTICS AT THE SCHOOL LEVEL*

In Model 4, presented above, school effects are conceived as the overall impact of the school for an "average" student. However, school effects can vary across students as schools may be more effective for one group of students than for another group. In this study, differential school effectiveness is assessed in the context of a growth model by adding parameters to the random-effects specification of Model 4 to allow the coefficients for the student-level variables *Gender* and *SES* to vary randomly at level 3 (school level), as suggested by Jesson and Gray (1991). Equation 9 shows the adittion of a school-level random effect for the variable *Gender*. Due to the complex variance structure of the model, it was not possible to test differential effects for all possible pupil

[34] This variable is a composite indicator, created and calculated by the Chilean Ministry of Education in the context of the SIMCE Assessment System. The variables considered in this indicator are: Mother's level of education, Father's level of education, Average monthly household income and the Social risk index of the school. These variables were standardised before collapsing them into one. The final variable was then categorised in five ordered levels, defined from 1 to 5, with level 1 being the lowest socio-economic level.

groupings within a single model. Thus, an equivalent model is separately fitted for the student variable *SES*.

EQUATION 9

$$Y_{tij} = \beta_{0ij} + \beta_{1ij}t_{tij} + \beta_2 t_{tij}^2 + e_{tij},$$

$$\beta_{0ij} = \beta_{00j} + \beta_{010}Cohort2_{ij} + \beta_{020}Cohort3_{ij} + \beta_{030}Cohort4_{ij} + \beta_{20j}Female_{ij} + \beta_{050}SES_{ij} + \beta_{060}Books_{ij} + \beta_{070}Age_{ij} + u_{0ij}$$

$$\beta_{1ij} = \beta_{10j} + \beta_{110}Cohort2_{ij} + \beta_{120}Cohort3_{ij} + \beta_{130}Cohort4_{ij} + \beta_{140}Female_{ij} + \beta_{150}SES_{ij} + \beta_{160}Books_{ij} + \beta_{170}Age_{ij} + u_{1ij}$$

$$\beta_{00j} = \beta_{000} + \beta_{200}Achievement\ Mean_j + \beta_{300}Achievement\ SD_j + \beta_{400}SES_j + u_{00j}$$

$$\beta_{10j} = \beta_{100} + \beta_{500}Achievement\ Mean_j + \beta_{600}Achievement\ SD_j + \beta_{700}SES_j + u_{01j}$$

$$\beta_{20j} = \beta_{040}Female_{ij} + u_{02j}$$

$$\begin{pmatrix} u_{00j} \\ u_{01j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u00}^2 & & \\ \sigma_{u00u01} & \sigma_{u01}^2 & \\ \sigma_{u00u02} & \sigma_{u01u02} & \sigma_{u02}^2 \end{pmatrix}\right]$$

$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]$$

$$e_{tij} \sim N(0, \sigma_e^2).$$

5.4 RESULTS

The results section is structured as follows. First, the variation on outcomes at the school level is analysed based on the results from Models 3 and 4. Secondly, school-level predictors are examined, with a focus on school compositional effects. Then, the relationship between language and mathematics school residuals is explored. Finally, school differential effects across groups are tested.

5.4.1 *THE MAGNITUDE OF SCHOOL EFFECTS*

Results from Model 3 in Chapter 4 show a substantive variation on achievement status across schools. More interestingly, schools also contribute to differences in growth rates (see random effects in Table 12). To get an impression of these figures, one has to bear in mind that the square root of the variance results is the standard deviation and that the components reflect the variation in the size of the regression coefficients. At the school level, the prediction equation that captures 68% of the schools in Cohort 1, in language, fluctuates between

$$\boldsymbol{Y = 347.750 + (34.582 \pm 3.124)Time + -7.728Time^2},$$

and, in mathematics, it fluctuates between

$$\boldsymbol{Y = 344.462 + (31.056 \pm 3.041)Time + -7.995Time^2},$$

indicating large differences in growth across schools. Indeed, most of the variance in growth appears to be at the school level: 76.9% for the linear component in language and 94.0% for the linear component in maths. This yields particularly large *d*-type effect sizes (0.877 and 0.970 for language and maths growth, respectively[35]).

In the field of educational effectiveness research, the development of three-level growth models has elicited a debate on whether schools have a larger effect on student status (intercept) than on student growth (slope) (Anumendem et al., 2013; De Fraine, Damme and Onghena, 2007; De Fraine et al., 2005; May, Supovitz and Perda, 2004; Raudenbush, 1989). Previous research has usually found that schools have a larger impact on their students' growth than on their students' outcomes at a certain point in time. For example, Raudenbush (1989, 1995) found that the school effect on student initial status for mathematics was 14% whereas the school effect on the learning rates was over 80%.

Here, in both language and maths, the school effects on the linear growth rate are larger than the school effects on students' achievement status (intercept), which are 34.4% and 41,9%, respectively. Also, school effects on students' achievement status (intercept) are similar to those found in previous studies in Chile using cross-sectional data which have estimated a between-school variance ranging from 30 to 50% (Mizala, Romaguera and Ostoic, 2004; OECD, 2013a; Willms and Somer, 2001). These results confirm what has been demonstrated previously in the literature: that studies that measure school effects by looking at student achievement cross-sectionally or longitudinally but over a short

---

[35] The *d*-type effect size is calculated as proposed by Rowan et al. (2002). Thus, for Model 2: $= \frac{\sqrt{\sigma^2_{u01}}}{\sqrt{(\sigma^2_{u1} + \sigma^2_{u01})}}$.

period of time (e.g., over two years) tend to underestimate the effect of the school (Kyriakides and Creemers, 2008; Rowan, Correnti and Miller, 2002).

These findings add to the evidence that longitudinal studies examining student growth are more likely to demonstrate school effects of larger magnitude than studies using covariate-adjustment and gain models (Raudenbush, 1995; Teddlie and Reynolds, 2000). In fact, the sizes of the school effects found here are somewhat larger to those found in previous studies using alike model specifications for similar outcomes (e.g., Guldemond and Bosker, 2009; Lenkeit, 2012; Palardy, 2010; Rowan, Correnti and Miller, 2002). This is also in line with previous research suggesting that higher school effects are found in emerging economies when compared to post-industrialised countries (Scheerens, 2001; Willms and Somer, 2001).

However, school effects, as specified in Model 3, should not be interpreted as a measure of school effectiveness as it is very likely that an important proportion of these effects are due to school intake characteristics. It becomes relevant then to identify those variables that could explain variance at the school levels by means of incorporating them into the model as predictors of the intercept and linear factors. This, in turn, would constitute a contextualised value-added model. The results of these analyses are presented in the following section.

As shown in Table 13, the variance components on achievement status and growth rates, both at the student and school level, remain sizeable once school variables are introduced in Model 4. Most of the variance in growth still appears to be at the school level (i.e., 74.1% and 93.9% for the linear component in language and mathematics, respectively). This still yields large *d*-type effect sizes of d = 0.861 and d = 0.969 for achievement growth in language and mathematics, respectively. In both language and maths, the school effects on the linear growth rate continue to be larger than the school effects on students' achievement status (intercept), which are 8.8% and 10.2%, respectively. In addition, school effects on achievement growth are larger for mathematics than for language. This result is in line with previous research suggesting that school effects are larger in subject areas that are typically learned largely at school, as with mathematics, where exposure is limited in the family and the community (Teddlie and Reynolds, 2000; Thomas et al., 1997a).

Figure 24 plots the school lines predicted by Model 4. These graphs depict the variation in schools' mean achievement and achievement growth as the intercepts and slopes of these regression lines differ substantially. The correlations between the random intercepts (the average proficiency in Grade 3 in a school) and random slopes at the school level are negative and of high to moderate size in language and mathematics (-0.549 and -0.541, respectively). Figure 24 also shows a subtle fanning-in pattern for the school prediction lines that is implied by this negative intercept/slope covariance at the school level.

Model 4, as well as all of the models in this study, has been fitted by coding the first of the three time points available as zero. It has been suggested in the literature that time coding might affect the size of the school effect in the sense that school effect on student status will be larger when the intercept refers to a later point in time (Biesanz et al., 2004; May, Supovitz and Perda, 2004). Figure 25 shows the school-level variance function for the starting, mid and last time points, based on Model 4. The trends for both subjects imply that the effect of the school is similar across time points, although somewhat smaller for the mid time point and relatively larger in the last time point. Thus, choosing the first time point as the reference seems a conservative approach.

TABLE 13: FULL MODEL 4 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

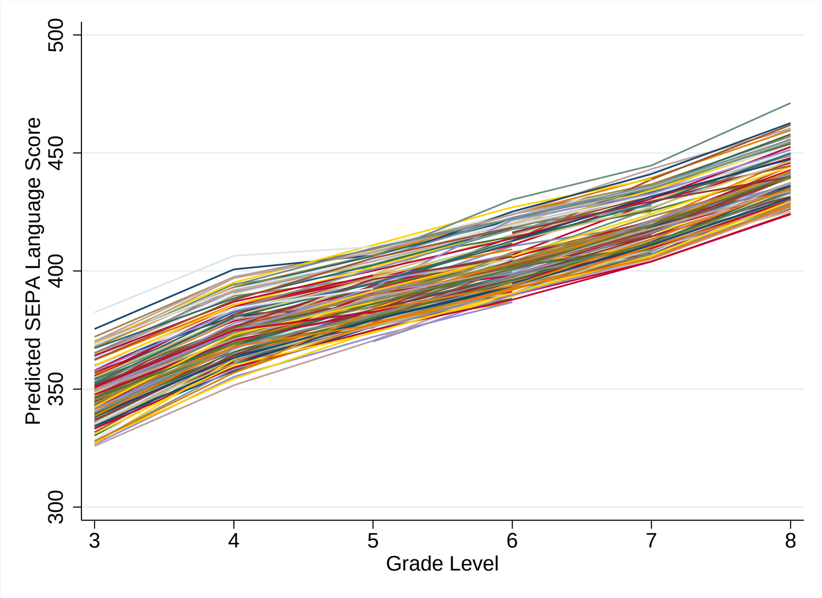
	MODEL 4 LANGUAGE		MODEL 4 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Fixed Part				
Intercept	348.363 (0.544)	347.276 - 349.450	345.400 (0.603)	344.146 - 346.653
Time	34.511 (1.087)	31.822 - 37.200	30.994 (1.945)	25.799 - 36.189
Time2	-7.725 (0.573)	-9.211 - -6.238	-7.995 (1.091)	-10.968 - -5.023
Cohort 2	25.074 (0.390)	24.308 - 25.840	23.083 (0.480)	22.075 - 24.091
Cohort 3	36.849 (0.516)	35.786 - 37.913	33.598 (0.426)	32.749 - 34.447
Cohort 4	54.766 (0.433)	53.910 - 55.623	56.152 (0.476)	55.171 - 57.134
Time x Cohort 2	-10.032 (0.794)	-11.737 - -8.326	-16.771 (1.967)	-21.946 - -11.596
Time x Cohort 3	-10.018 (1.088)	-12.551 - -7.485	-6.033 (2.184)	-11.808 - -0.259
Time x Cohort 4	-13.948 (1.006)	-16.252 - -11.644	-4.650 (2.373)	-10.979 - 1.679
Time2 x Cohort 2	3.076 (0.375)	2.272 - 3.880	8.679 (0.994)	6.048 - 11.310
Time2 x Cohort 3	2.878 (0.523)	1.655 - 4.102	5.981 (1.146)	2.924 - 9.038
Time2 x Cohort 4	6.434 (0.514)	5.234 - 7.634	6.431 (1.220)	3.157 - 9.704
Female student	2.019 (0.309)	1.400 - 2.638	-2.004 (0.315)	-2.652 - -1.356
Age	-1.696 (0.268)	-2.227 - -1.165	-1.643 (0.281)	-2.217 - -1.070
SES	0.684 (0.307)	-0.036 - 1.404	1.116 (0.307)	0.386 - 1.846
Number books at home	2.366 (0.152)	2.067 - 2.664	2.021 (0.164)	1.689 - 2.354
Time x Female	0.366 (0.203)	-0.090 - 0.822	-0.697 (0.152)	-1.012 - -0.382
Time x Age	-0.348 (0.134)	-0.620 - -0.076	-0.475 (0.209)	-0.967 - 0.017
Time x SES	-0.045 (0.285)	-0.760 - 0.670	-0.289 (0.140)	-0.580 - 0.002
Time x Number books at home	-0.212 (0.109)	-0.455 - 0.031	-0.088 (0.083)	-0.259 - 0.084
Achievement Mean	0.213 (0.033)	0.149 - 0.277	0.270 (0.029)	0.210 - 0.329
Achievement SD	-0.173 (0.124)	-0.418 - 0.071	-0.179 (0.098)	-0.372 - 0.013
School SES	2.532 (0.744)	0.989 - 4.075	0.432 (0.757)	-1.134 - 1.998
Time x Achievement Mean	-0.050 (0.025)	-0.103 - 0.003	-0.039 (0.026)	-0.097 - 0.018
Time x Achievement SD	-0.150 (0.102)	-0.370 - 0.069	-0.033 (0.069)	-0.170 - 0.104
Time x School SES	-0.277 (0.451)	-1.205 - 0.651	0.818 (0.633)	-0.597 - 2.233
Random Part				

	MODEL 4 LANGUAGE		MODEL 4 MATHEMATICS	
	M (SE)	**95% CI**	**M (SE)**	**95% CI**
Random Level 3 (School)				
Intercept	19.493 (3.462)	12.384 - 26.601	18.778 (4.365)	9.017 - 28.538
Time/Intercept	-6.907 (1.613)	-10.151 - -3.663	-6.972 (2.005)	-11.270 - -2.674
Time	8.131 (1.744)	4.244 - 12.018	8.838 (1.349)	6.120 - 11.556
Random Level 2 (Student)				
Intercept	202.067 (3.463)	195.234 - 208.900	166.145 (3.077)	159.935 - 172.355
Time/Intercept	-9.100 (1.619)	-12.46 - -5.739	-3.173 (0.928)	-5.052 - -1.295
Time	2.844 (0.921)	1.027 - 4.660	0.575 (0.174)	0.234 - 0.917
Random Level 1 (Occasion)				
Intercept	119.201 (2.067)	114.524 - 123.877	113.732 (1.237)	111.066 - 116.399
School Initial Status – Growth Correlation	-0.549		-0.541	
Student Initial Status – Growth Correlation	-0.380		-0.325	
School ICC Initial Status	0.088		0.102	
School ICC Growth	0.741		0.939	
DIC	**467,891**		**463,208**	
pD	**17,198**		**16,373**	
Units: Schools	**156**		**156**	
Units: Students	**19,704**		**19,704**	
Units: Occasions	**59,112**		**59,112**	

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions, student and school levels. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

FIGURE 24: SCHOOL LINES BASED ON MODEL 4

LANGUAGE



MATHEMATICS

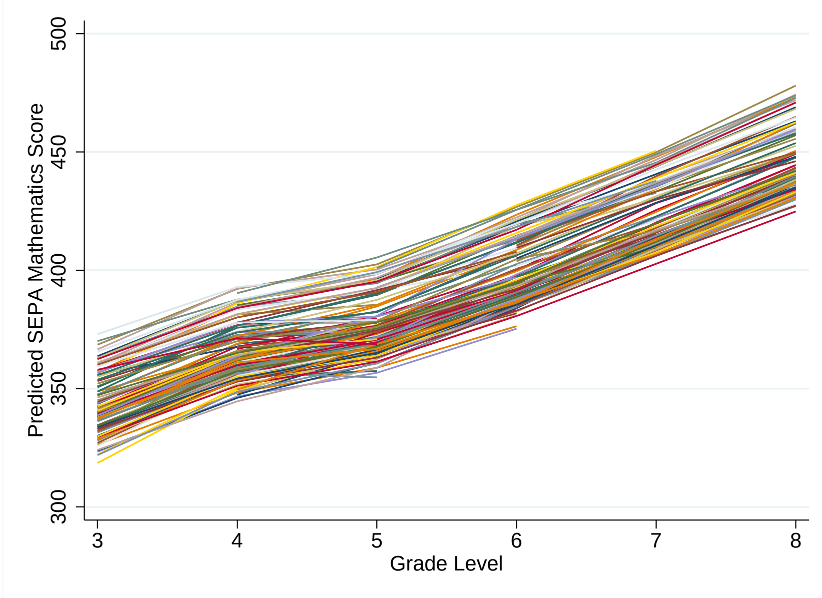
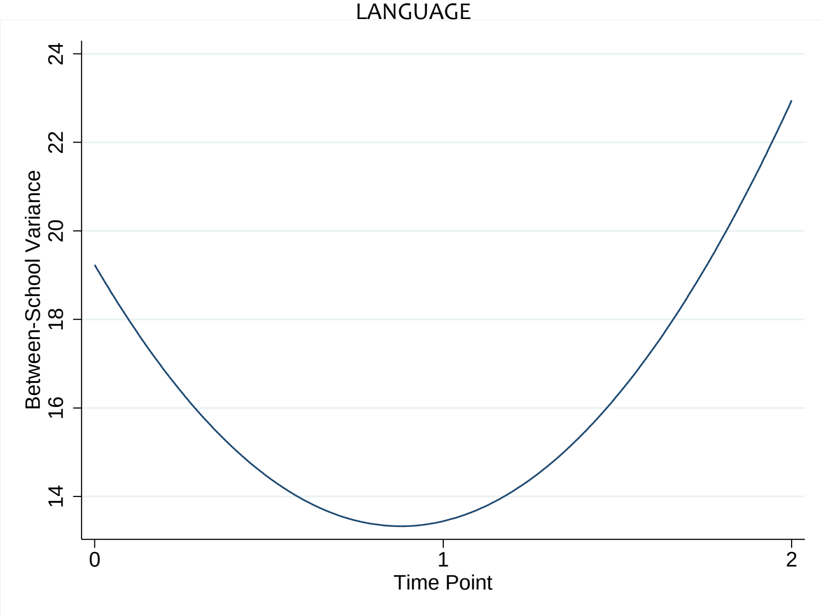


FIGURE 25: SCHOOL-LEVEL VARIANCE FUNCTION

LANGUAGE



MATHEMATICS

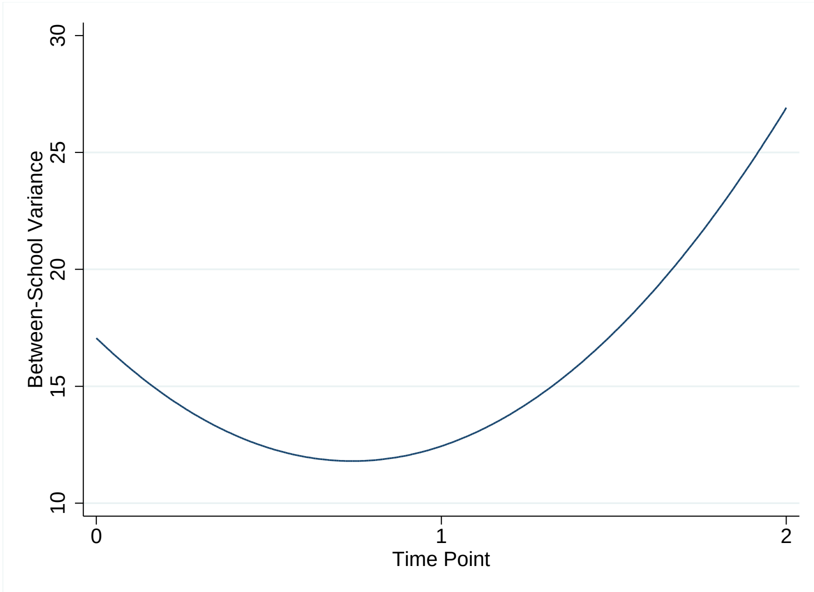


FIGURE 26: SCHOOL RESIDUALS FOR LANGUAGE BASED ON MODEL 4

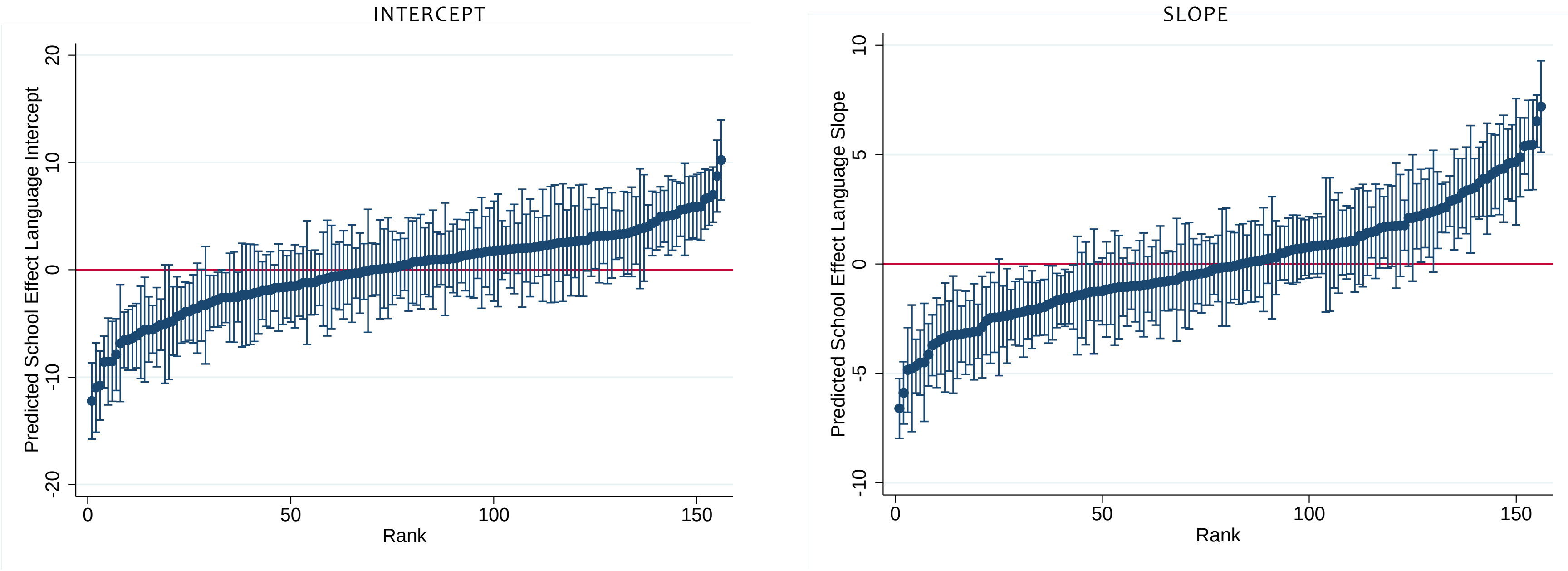
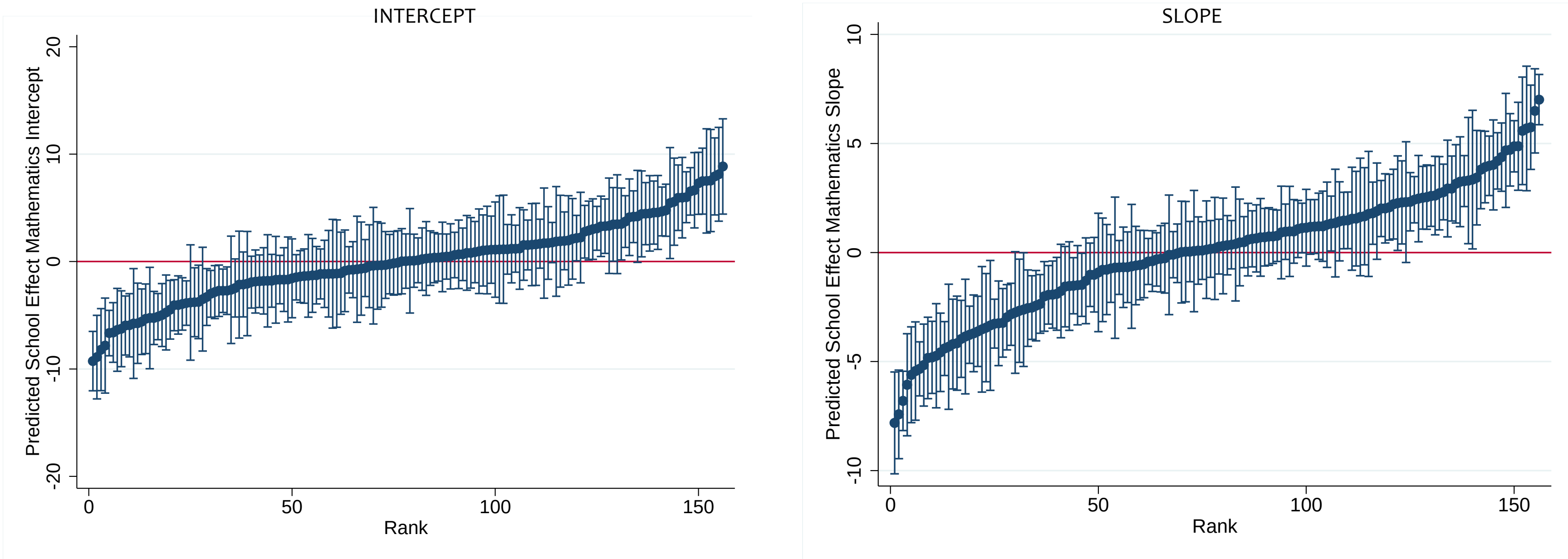


FIGURE 27: SCHOOL RESIDUALS FOR MATHEMATICS BASED ON MODEL 4



To illustrate the magnitude of school effects, the 156 level-3 residuals are plotted in Figures 26 and 27, one for each school in the data set. These caterpillar plots graph each residual, obtained from Model 4, against their rank order, accompanied by error bars corresponding to confidence intervals.

School effects on achievement growth (slope) are estimated with greater precision than school effects on achievement status (intercept). Nonetheless, it is possible to distinguish several outliers at both ends of each graph. The results show that there is greater variation in school effects for mathematics than for language. In language, the confidence intervals of these residuals do not overlap zero for a group of 53 and 76 schools at the lower and upper end of the intercept and residual plots, respectively. In mathematics, the number of outlier schools is 65 for the intercept and 83 for the slope residuals. This means that more than 34% of the schools differ significantly from the average school effect at the 5% level.

In Table 14, schools have been classified according to their slope residuals for the two outcome measures. If the confidence intervals for a school residual do not overlap zero, the value added residual is significantly different either above or below expectation and identified as an outlier. The table shows that schools that are effective in one subject tend, to some extent, to be effective in the other subject. The consistency of school effects is analysed in greater detail later in this chapter.

TABLE 14: CROSS TABULATION OF SCHOOL EFFECTS ON STUDENT ACHIEVEMENT GROWTH

	LANGUAGE			
MATHEMATICS	**Significant ++**	**Non-Significant +**	**Non-Significant -**	**Significant - -**
Significant ++	16 (10.26%)	12 (7.69%)	9 (5.77%)	5 (3.21%)
Non-Significant +	10 (6.41%)	10 (6.41%)	11 (7.05%)	14 (8.97%)
Non-Significant -	5 (3.21%)	11 (7.05%)	6 (3.85%)	6 (3.85%)
Significant - -	2 (1.28%)	7 (4.49%)	14 (8.97%)	18 (11.54%)

Note: Total number of schools = 156. Percentages given in brackets.

Estimated residuals specified in Model 4 are also used to check an important model assumption: that the residuals at each level follow normal distributions. Appendix 7 shows the normal probability plots, where ranked school residual are graphed against corresponding points on a normal distribution curve. The plots look fairly linear, which suggests that the assumption of normality is met.

5.4.2 *COMPOSITIONAL EFFECTS*

The results presented so far indicate significant variation among schools, both in terms of intercepts and rates of achievement, encouraging a search for differences in school characteristics that might account for this variation. A series of preliminary models were fitted to the data to test the effect of several school variables as predictors of achievement status and rates of change but, in the interest of space, only the final model is presented. In these alternative models, the additional variables tested were *School Size* (number of students enrolled in the school), *Teacher/Student Ratio*, *School Geographical Location* (urban/rural) and *School Sector* (Private Non-Subsidised/Private Subsidised/Public)[36]. These covariates were excluded from Model 4 as they were not found to predict student achievement status nor growth after controlling for the rest of the student and school variables and their inclusion did not improve model fit. In what follows, fixed effect results from Model 4, which incorporates a selection of school-level predictors, are presented.

In Model 4, the school variables of interest are *Achievement Mean*, *Achievement SD* and *School SES*. Including these variables led to an improvement in model fit, as shown by the reduction of the DIC value, for both subjects, in comparison to Model 3 (Δ*DIC* = -9 in language, and Δ*DIC* = -11 in mathematics). The addition of the school- and student-level variables explained a total of 86.70% and 88.29% of the school-level variance in attainment observed in Model 2, for language and mathematics, respectively. As shown in Table 13, school variables that were found to be associated with achievement status were *Achievement Mean* and *School SES*, although the latter only in language. This suggests that it is academically beneficial for students to be part of schools with high achievement levels. On the other hand, the variance in achievement within schools does

[36] Variables capturing schools' environment or organizational practices were not available in the sources of secondary data.

not appear to have a systematic effect on student achievement. The evidence that school SES is positively related to the academic achievement of students in language, even after accounting for family SES, supports the hypothesis of double jeopardy. None of the school-level variables tested were substantially related to achievement growth.

Interestingly, other longitudinal studies investigating school composition have also demonstrated significant school composition effects on achievement on the first measurement occasion and no, or considerably smaller, on learning growth (Belfi et al., 2013; Guldemond and Bosker, 2009; Luyten, Schildkamp and Folmer, 2009). In Figures 28 to 30, these trends are depicted, showing the differences in the relative positions of students in schools with different SES levels, and in different school overall achievement and achievement SD quartiles, and how they tend to remain stable across time.

An important stream of studies on the Chilean system focuses on differences between school sectors and, in particular, on the relative effectiveness of private-voucher versus public schools (e.g., Anand, Mizala and Repetto, 2009; Lara, Mizala and Repetto, 2011; McEwan and Carnoy, 2000; Mizala and Romaguera, 2000; Sapelli and Vial, 2002; Tokman, 2002). This distinction is highly relevant to the Chilean context as it emerges from major policy reforms carried out in the country during the last decades in the line of promoting the participation of the private subsidised sector. Alternative models explored showed that being at a private non-subsidised or at a private subsidised school, in comparison to attending a public school, was not associated with differences in student achievement growth, after controlling for the student and school variables in Model 4. Figure 31 shows the predicted growth trajectories for the different school sectors based on Model 4. These results are in line with previous research in Chile, in that they indicate that differences in school effectiveness measures between public and private subsidised schools (voucher schools) are negligible.

The evidence of marked variation in attainment and growth rates between schools even after controlling for relevant student characteristics and school intake, and the fact that the variation in achievement status and growth rates could not be fully explained by these variables, give room to the hypothesis of potential teacher effects. These effects will be explored in the following chapter.

FIGURE 28: GROWTH CURVES BY SCHOOL ACHIEVEMENT MEAN QUARTILES BASED ON MODEL 4

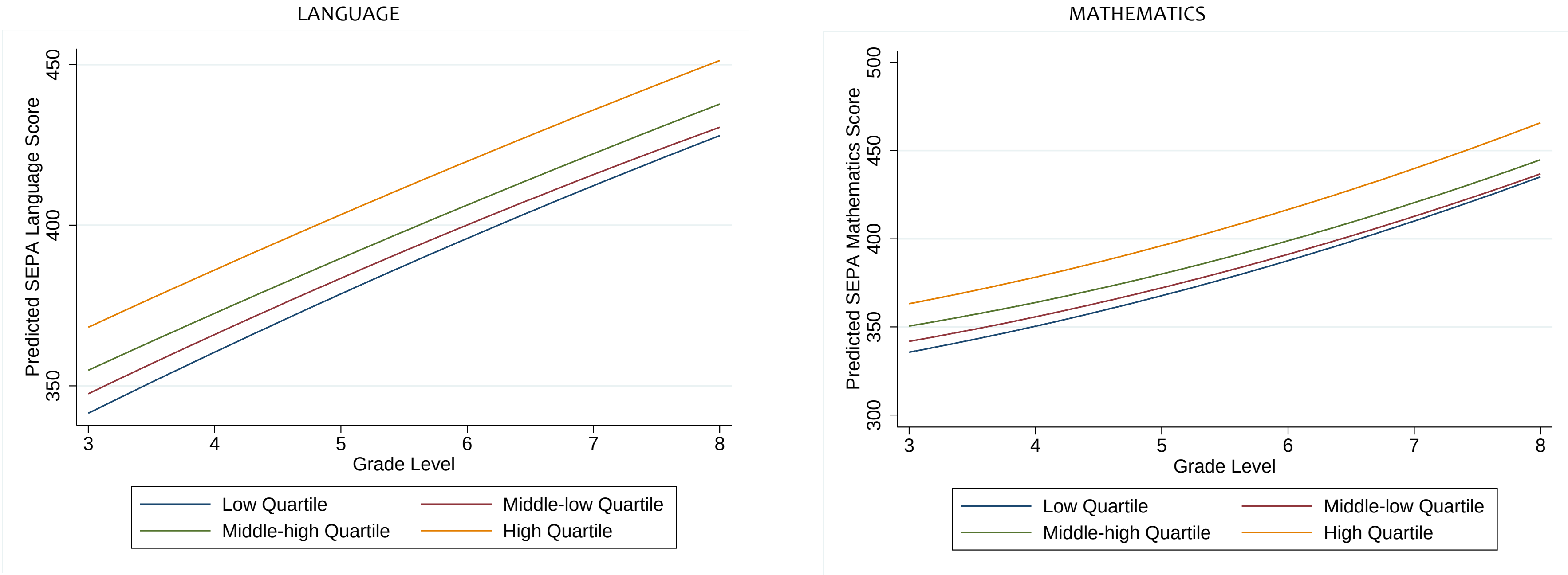


FIGURE 29: GROWTH CURVES BY SCHOOL ACHIEVEMENT SD QUARTILES BASED ON MODEL 4

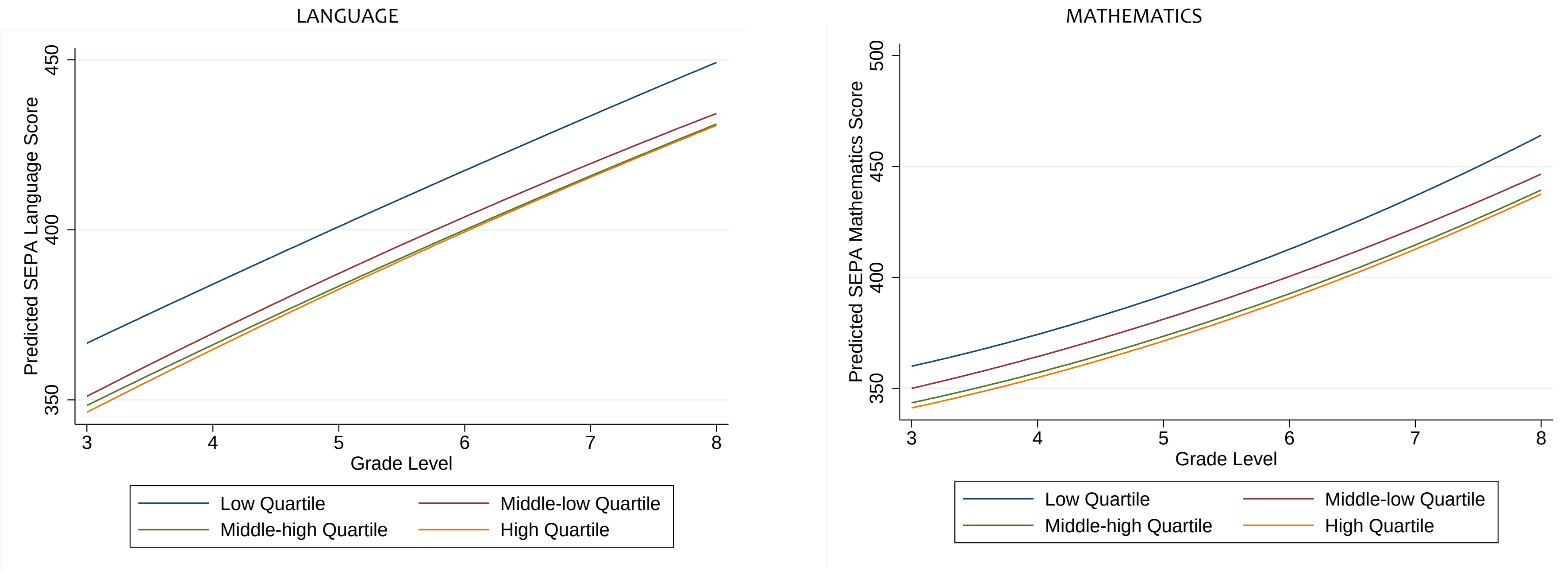


FIGURE 30: GROWTH CURVES BY SCHOOL SES CATEGORIES BASED ON MODEL 4

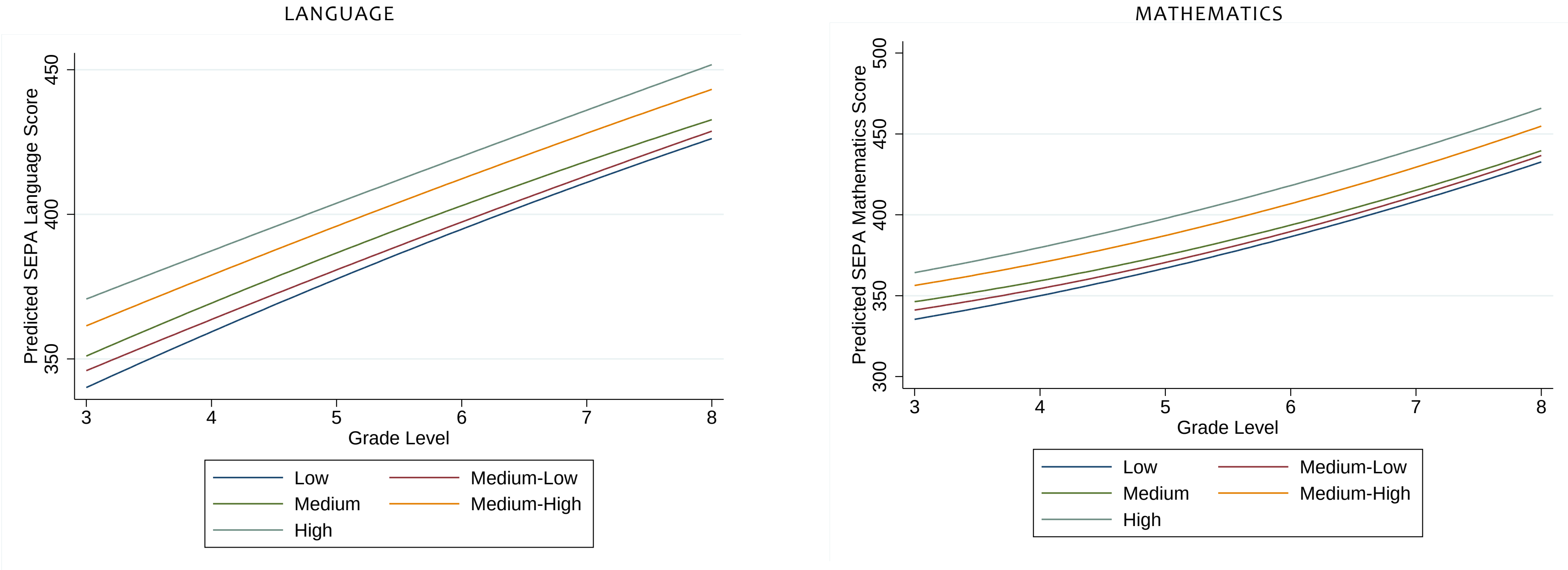
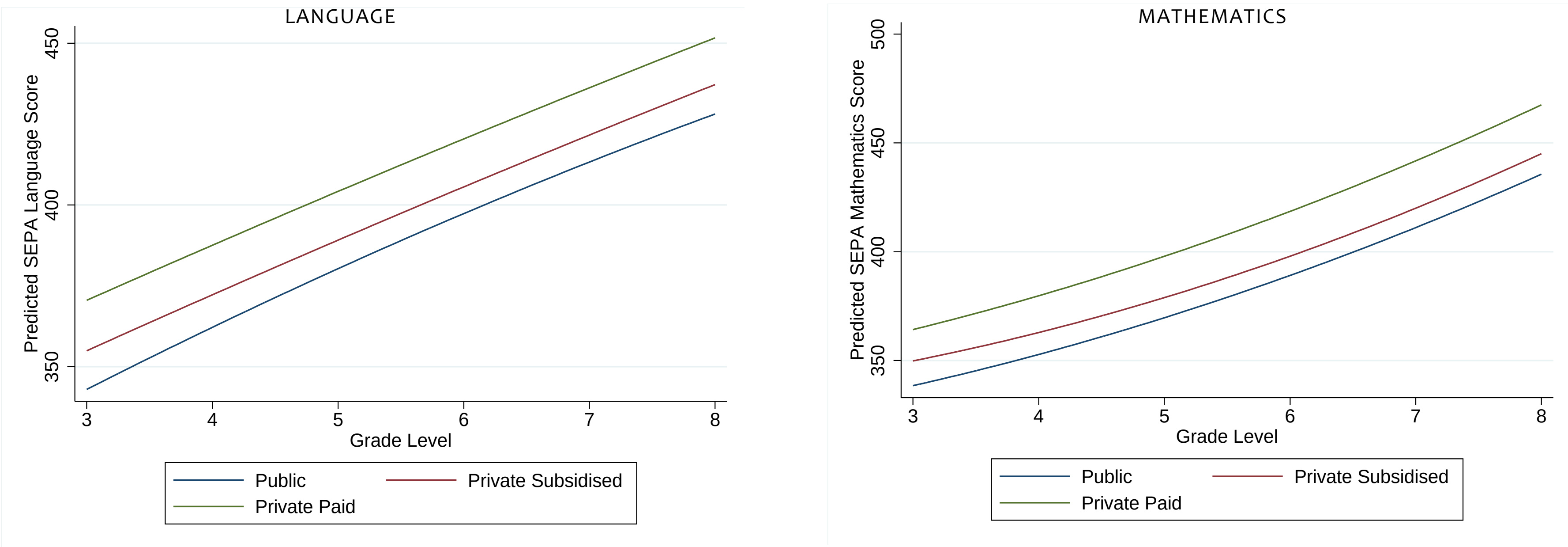


FIGURE 31: GROWTH CURVES BY SCHOOL SECTOR BASED ON MODEL 4



5.4.3 *The Consistency of School Effects across Subjects*

In order to assess whether schools have consistent effects on students' achievement trajectories across academic subjects, correlation coefficients between school residuals in language and mathematics are calculated for the 156 schools in our sample. As shown in Table 15, the moderate correlation between slope residuals (r = 0.425) suggests some degree of coherence on school effectiveness across academic subjects. All in all, schools that are effective in language are, to some extent, also effective in mathematics. However, as these correlations are by no means perfect, it is possible to hypothesise that some schools might be better at promoting achievement in one of the two subjects. Scatterplots in Figure 32 also depict the relationship of school residuals across the academic subjects.

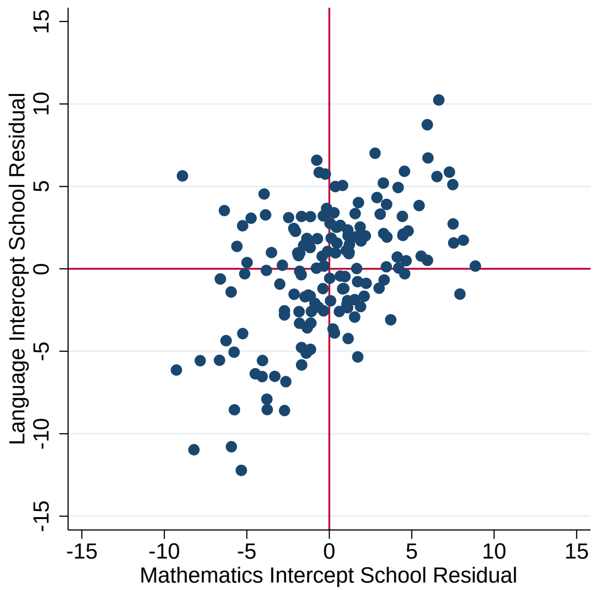
TABLE 15: CORRELATION BETWEEN MODEL 4 SCHOOL RESIDUALS

	Language Intercept Residual	Language Slope Residual	Mathematics Intercept Residual	Mathematics Slope Residual
Language Intercept Residual	1			
Language Slope Residual	-0.563***	1		
Mathematics Intercept Residual	0.489***	-0.261***	1	
Mathematics Slope Residual	-0.133	0.425***	-0.569***	1

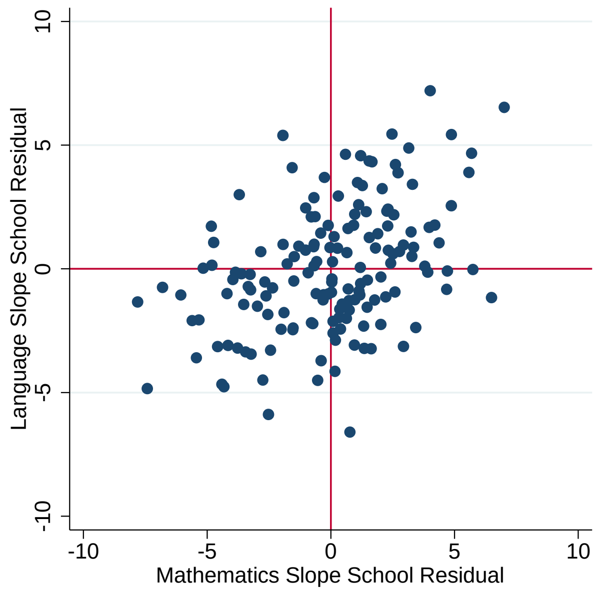
Note: Total number of schools = 156. Pearson correlation was used. Significance of parameter estimates is indicated by asterisks: *** = p<.001.

FIGURE 32: SCATTERPLOTS OF LANGUAGE AND MATHEMATICS SCHOOL RESIDUALS BASED ON MODEL 4

INTERCEPT



SLOPE



5.4.4 *Differential Effects of Schools across Student Groups*

A fourth question regarding school effects is whether schools have consistent effects across student groups. Differential school effects upon students with different characteristics are considered to be an equity issue (Levine and Lezotte, 1990). To investigate this issue, the specification of Model 4, which assumes that the effects of student-level variables on achievement are the same in all schools, is changed to that of Model 5, where the effects of student *Gender* and *SES* on achievement growth vary randomly across schools. The effect of only one of these independent variables is allowed to vary randomly in this way in each version of Model 5.

In Table 16, the fit of Model 4 to the data is compared to that of the different versions of Model 5, on the basis of the deviance information criterion (DIC). In language, judging by the differences of more than 5 units in DIC values between Models 4 and 5, in favour of the latter, the analyses show that the student variables *Gender* and *SES* have different effects on achievement growth across schools.

TABLE 16: DEVIANCE INFORMATION CRITERION (DIC) FOR MODELS 4 AND 5

	DIC		
RANDOM EFFECT	**MODEL 4**		**MODEL 5**
LANGUAGE			
Gender	467,891	>	467,884
SES		>	467,530
MATHEMATICS			
Gender	463,208	=	463,208
SES		>	463,020

These results are confirmed by the random effects results for these variables shown in Table 17, suggesting differences in schools' slopes. The approach proposed by Lockwood and McCaffrey (2009)[37] is used to calculate the magnitude of the interaction terms, and indicates that the fractions of the total school effect variance on language achievement growth that are due to the interactions with student gender and SES are 21 and 27%, respectively. In mathematics, in turn, *SES* has different effects on achievement growth across schools (29% of the total school variance on growth), confirming that students

[37] For Model 5 this is calculated as $\frac{\sigma^2_{u02}}{\sigma^2_{u01}+\sigma^2_{u02}}$.

from different socio-economic backgrounds do not progress equally across schools. The influence of gender in mathematics achievement growth does differ somewhat between schools, but these differences are smaller (17% of the total school variance on achievement growth).

TABLE 17: FULL MODEL 5 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 5 - DIFFERENTIAL SCHOOL EFFECT BY STUDENT GENDER				MODEL 5 - DIFFERENTIAL SCHOOL EFFECT BY STUDENT SES			
	LANGUAGE		MATHEMATICS		LANGUAGE		MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI
Fixed Part								
Intercept	348.43 (0.544)	347.343 - 349.517	345.417 (0.607)	344.155 - 346.679	348.343 (0.562)	347.204 - 349.483	345.439 (0.564)	344.288 - 346.591
Time	34.517 (1.087)	31.825 - 37.209	30.996 (1.944)	25.802 - 36.191	34.600 (1.107)	31.857 - 37.342	31.002 (1.872)	26.020 - 35.984
Time2	-7.726 (0.574)	-9.213 - -6.240	-7.996 (1.091)	-10.969 - -5.024	-7.764 (0.587)	-9.285 - -6.243	-8.010 (1.060)	-10.891 - -5.130
Cohort 2	25.086 (0.392)	24.315 - 25.857	23.080 (0.481)	22.070 - 24.090	25.151 (0.394)	24.376 - 25.926	23.104 (0.471)	22.123 - 24.085
Cohort 3	36.855 (0.517)	35.789 - 37.920	33.594 (0.425)	32.747 - 34.440	36.876 (0.520)	35.803 - 37.950	33.603 (0.420)	32.769 - 34.438
Cohort 4	54.758 (0.438)	53.891 - 55.624	56.132 (0.482)	55.140 - 57.125	54.802 (0.438)	53.936 - 55.669	56.210 (0.459)	55.278 - 57.142
Time x Cohort 2	-10.034 (0.795)	-11.740 - -8.328	-16.773 (1.967)	-21.947 - -11.599	-10.249 (0.834)	-12.063 - -8.436	-16.891 (1.941)	-21.985 - -11.797
Time x Cohort 3	-10.024 (1.087)	-12.556 - -7.492	-6.037 (2.182)	-11.809 - -0.265	-10.148 (1.087)	-12.674 - -7.621	-6.058 (2.142)	-11.710 - -0.407
Time x Cohort 4	-13.953 (1.008)	-16.258 - -11.647	-4.653 (2.373)	-10.981 - 1.674	-14.188 (0.996)	-16.451 - -11.925	-4.794 (2.327)	-10.985 - 1.397
Time2 x Cohort 2	3.078 (0.376)	2.273 - 3.882	8.680 (0.994)	6.049 - 11.311	3.164 (0.390)	2.321 - 4.008	8.741 (0.979)	6.155 - 11.327
Time2 x Cohort 3	2.881 (0.523)	1.658 - 4.105	5.984 (1.146)	2.927 - 9.041	2.935 (0.529)	1.696 - 4.174	6.009 (1.119)	3.032 - 8.985
Time2 x Cohort 4	6.436 (0.514)	5.236 - 7.636	6.433 (1.220)	3.159 - 9.706	6.542 (0.505)	5.372 - 7.712	6.503 (1.196)	3.301 - 9.705
Female student	1.897 (0.350)	1.197 - 2.598	-2.012 (0.343)	-2.710 - -1.314	2.010 (0.312)	1.382 - 2.638	-2.000 (0.312)	-2.642 - -1.357
Age	-1.699 (0.268)	-2.231 - -1.166	-1.640 (0.280)	-2.212 - -1.068	-1.700 (0.271)	-2.239 - -1.160	-1.653 (0.277)	-2.218 - -1.088
SES	0.686 (0.307)	-0.035 - 1.407	1.116 (0.307)	0.387 - 1.846	0.791 (0.319)	0.114 - 1.468	1.168 (0.333)	0.455 - 1.881
Number books at home	2.367 (0.151)	2.070 - 2.665	2.020 (0.163)	1.689 - 2.351	2.344 (0.150)	2.050 - 2.638	1.996 (0.167)	1.656 - 2.336
Time x Female	0.365 (0.202)	-0.090 - 0.820	-0.698 (0.152)	-1.013 - -0.382	0.368 (0.202)	-0.087 - 0.823	-0.701 (0.150)	-1.013 - -0.389
Time x Age	-0.348 (0.134)	-0.620 - -0.075	-0.476 (0.209)	-0.968 - 0.015	-0.340 (0.135)	-0.613 - -0.067	-0.465 (0.209)	-0.955 - 0.026
Time x SES	-0.046 (0.285)	-0.761 - 0.669	-0.290 (0.139)	-0.579 - 0.000	-0.070 (0.291)	-0.801 - 0.661	-0.296 (0.137)	-0.580 - -0.012
Time x Number books at home	-0.212 (0.109)	-0.455 - 0.032	-0.088 (0.083)	-0.259 - 0.084	-0.206 (0.112)	-0.457 - 0.045	-0.080 (0.085)	-0.256 - 0.097
Achievement Mean	0.213 (0.033)	0.148 - 0.279	0.269 (0.029)	0.210 - 0.328	0.231 (0.034)	0.164 - 0.297	0.276 (0.026)	0.226 - 0.327
Achievement SD	-0.176 (0.123)	-0.419 - 0.068	-0.186 (0.098)	-0.380 - 0.008	-0.150 (0.127)	-0.404 - 0.105	-0.162 (0.094)	-0.347 - 0.023
School SES	2.560 (0.779)	0.941 - 4.180	0.445 (0.773)	-1.153 - 2.043	2.280 (0.713)	0.828 - 3.731	0.361 (0.682)	-1.003 - 1.725
Time x Achievement Mean	-0.050 (0.025)	-0.103 - 0.003	-0.040 (0.026)	-0.098 - 0.018	-0.055 (0.025)	-0.108 - -0.003	-0.040 (0.026)	-0.098 - 0.017
Time x Achievement SD	-0.150 (0.103)	-0.373 - 0.072	-0.033 (0.071)	-0.175 - 0.109	-0.163 (0.101)	-0.380 - 0.054	-0.037 (0.071)	-0.179 - 0.104
Time x School SES	-0.230 (0.453)	-1.168 - 0.708	0.823 (0.628)	-0.583 - 2.229	-0.205 (0.437)	-1.099 - 0.688	0.842 (0.631)	-0.575 - 2.258

	MODEL 5 - DIFFERENTIAL SCHOOL EFFECT BY STUDENT GENDER				MODEL 5 - DIFFERENTIAL SCHOOL EFFECT BY STUDENT SES			
	LANGUAGE		MATHEMATICS		LANGUAGE		MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI	M (SE)	95% CI
Random Part								
Random Level 3 (School)								
Intercept	20.259 (4.024)	11.939 - 28.579	20.045 (4.790)	9.453 - 30.636	18.983 (3.567)	11.612 - 26.355	19.492 (5.020)	8.011 - 30.973
Time/Intercept	-6.462 (1.690)	-9.869 - -3.054	-7.036 (2.174)	-11.728 - -2.345	-7.184 (1.660)	-10.527 - -3.842	-7.261 (2.185)	-12.009 - -2.514
Time	8.169 (1.744)	4.289 - 12.049	8.872 (1.346)	6.162 - 11.582	8.385 (1.574)	5.030 - 11.740	9.001 (1.422)	6.111 - 11.891
Intercept/ Student Variable (Gender/SES)	-1.432 (1.450)	-4.294 - 1.430	-1.716 (1.438)	-4.573 - 1.140	-1.713 (1.329)	-4.480 - 1.055	-3.387 (1.831)	-7.482 - 0.707
Time/Student Variable (Gender/SES)	-0.893 (0.894)	-2.657 - 0.871	0.124 (0.890)	-1.640 - 1.888	1.153 (0.937)	-0.852 - 3.159	1.265 (0.902)	-0.606 - 3.135
Student Variable (Gender/SES)	2.172 (0.819)	0.565 - 3.778	1.761 (0.714)	0.355 - 3.168	3.063 (1.033)	0.817 - 5.309	3.663 (1.101)	1.298 - 6.028
Random Level 2 (Student)								
Intercept	201.609 (3.419)	194.857 - 208.361	165.685 (3.074)	159.480 - 171.890	202.768 (3.423)	196.015 - 209.521	165.910 (3.222)	159.319 - 172.501
Time/Intercept	-9.127 (1.556)	-12.383 - -5.872	-3.114 (0.932)	-5.004 - -1.224	-9.661 (1.679)	-13.226 - -6.095	-3.178 (0.927)	-5.056 - -1.299
Time	2.922 (0.814)	1.304 - 4.539	0.568 (0.145)	0.284 - 0.852	3.416 (0.877)	1.673 - 5.159	0.654 (0.159)	0.341 - 0.966
Random Level 1 (Occasion)								
Intercept	119.129 (2.027)	114.468 - 123.791	113.744 (1.233)	111.085 - 116.404	118.149 (2.047)	113.464 - 122.834	113.090 (1.184)	110.570 - 115.610
DIC	467,884		463,208		467,530		463,020	
pD	17,209		16,370		17,541		16,471	
Units: Schools	**156**		**156**		**156**		**156**	
Units: Students	**19,704**		**19,704**		**19,704**		**19,704**	
Units: Occasions	**59,112**		**59,112**		**59,112**		**59,112**	

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions, student and school levels. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

5.5 Conclusions

The present chapter investigated school effects on the achievement growth trajectories of primary school students in Chile. The study was facilitated by the use of three-level longitudinal growth models in combination with an accelerated longitudinal design.

Results indicate that the cross-cohort performance of schools differed depending on whether the mean achievement status or growth of students was considered. Across the four cohorts studied, the relationship between the initial achievement status of students and students' subsequent achievement progress is negative as schools with high initial achievement are generally less likely than schools with low initial achievement to have high mean achievement growth.

School effects on students' growth trajectories are sizeable and moderately consistent across the two subjects. Compositional effects are present as school achievement mean (calculated using external sources and not aggregated from the dependant variable) predicted achievement status on both subjects. Also, in language, the school's SES composition was found to have effects on achievement outcomes over and above the student SES, supporting the double jeopardy hypothesis. While school composition was associated with achievement levels, these same factors were not statistically related to achievement growth rates. School sector was not a significant predictor of student achievement growth, after controlling for student background and school composition. This is an important finding in the Chilean context given claims that private subsidised schools are better than public schools. The results show that sector differences are due to differences in student intake and not to school effectiveness.

In general, both the student and school-level variables tested had substantive effects on students' achievement status but less so on growth in achievement. These results are supported by the literature (Rowan, Correnti and Miller, 2002) and suggest that, when the analysis shifts from concern with students' achievement status to a concern with students' growth in achievement, social background and demographic characteristics, as well as school composition become relatively insignificant predictors of academic development.

The findings also add to the evidence that longitudinal studies examining student growth are more likely to demonstrate school effects of larger magnitude than studies using covariate-adjustment and gain models and that school contributions are larger on student achievement growth than on achievement status. In fact, the sizes of school effects found here are somewhat larger than those found in previous studies using similar model specifications and outcomes in the other national contexts, and could indicate that larger school effects can be found in emerging economies when compared to post-industrialised countries.

The phenomenon of differential educational effects across student groups within schools was also studied in this chapter. In both subjects, schools were found to be differentially effective across student from different socioeconomic status. In language, schools were also found to be differentially effective based on student gender.

The fact that a sizeable percentage of between-school variance remains after adjustment for both student background variables and school compositional effects indicates that malleable educational conditions rather than merely student selection factors are likely to account for the remaining between-school variance. Studying teacher effects could help to disentangle these large school effects. All of the relevant features of students' achievement trajectories found in this chapter will be carried forward to the models introduced in the following chapter, in which the magnitude, consistency and predictors of teacher effects are explored in detail. In terms of statistical modelling, the forthcoming analyses imply the introduction of cross-classified and multiple membership models that allows the estimation of the contribution of teachers to students' achievement trajectories and the analysis of the fixed effect of a series of teacher-level variables.

CHAPTER 6: TEACHER EFFECTS ON STUDENT ACHIEVEMENT TRAJECTORIES

6.1 INTRODUCTION

As shown in the previous chapters, the use of three time points and statistical techniques well suited to model longitudinal data contribute to the methodological development of better approaches to measure student achievement trajectories. The multilevel approach applied to a three time point accelerated longitudinal design enables the separation of student and school variation from variation due to test-level measurement error while drawing on a substantial source of intra-individual variation to examine changes in student attainment over time. This allowed the estimation of the trajectories of the achievement gaps attributed to four different student characteristics: gender, age, SES and number of books at home from Grade levels 3 to 8 of primary school in Chile. In addition, the effects of schools on both student achievement status and growth were estimated.

The present chapter uses linked data from seven different secondary sources (see Chapter 3) and draws from the models fitted in Chapter 5 to estimate the overall size of teacher effects on student achievement trajectories in Chile and to investigate several properties of such effects. To do so, more sophisticated models, that incorporate cross-classification and multiple membership components, are introduced.

6.2 RESEARCH QUESTIONS

As presented in section 3.1, the third aim of this study is to investigate the magnitude of generic teacher effects in Chile, their consistency across subjects, their teacher-level predictors and whether they accumulate over time. This aim unfolds into the following four research questions, to be answered in this chapter:

1. How large are the effects of teachers on the achievement of their students and on their changes in achievement over time?
2. How consistent are teacher effects across different academic subjects (i.e., language and mathematics)?
3. What teacher characteristics account for the variation in student trajectories over time?
4. Do teacher effects accumulate over time?

6.3 Method of Analysis

6.3.1 *Data*

As in the previous chapters, our analytic sample consists of students in Grades 3 to 8 who took the SEPA language or mathematics tests. Analyses were carried out considering only those schools with 22 or more students and those teachers with at least 5 students in each of the three years assessed. After performing multiple imputation (details discussed in Chapter 3), the language and mathematics samples are balanced, that is, the number of time-point observations is three for each of the students. Table 18 shows the sample size in terms of number of occasions, students, teachers and schools, the classifications in the data[38]. Only the number of teachers is different between the language (N=851) and the mathematics (N=812) samples. As described in Chapter 3, these samples include a broad range of schools, teachers and students.

TABLE 18: SAMPLE SIZE BY CLASSIFICATION

Classification	N
Schools	156
Teachers	851 (Language) / 812 (Mathematics)
Students	19,704
Occasions	59,112

[38] Note that the concept of levels of units breaks down when we consider cross-classified data structures, which is the case in this chapter, and so some authors refer to the different groupings of units in the data structure as classifications.

Likewise the models fitted in the previous chapters, dependent variables are language and mathematics achievement. They are based on students' scores in the language and mathematics tests that are part of the SEPA test batteries. These standardised achievement tests were developed by a group of experts at the MIDE Assessment Centre of the Catholic University of Chile. Item Response Theory (IRT) was used to scale the language and mathematics tests across year levels and cohorts.

6.3.2 *Cross-Classified Data*

Up to now, the multilevel data structures modelled have been such that a lower-level unit (i.e., an occasion) has been perfectly nested in one (or more) higher-level unit (i.e., a student and a school). A limitation of the three-level hierarchical model in assessing contextual effects on individual change is that the model applies only to those persons who remain in a single context during the course of the investigation. When persons cross contextual boundaries during the study, the data no longer have a nested structure. Rather, the structure involves cross-classification of persons by social setting. Just such a migration occurs when one attempts to study the effects of teachers on children's cognitive growth (Raudenbush and Bryk, 2002). As we attempt to specify teacher effects, we are drawn to a situation in which students can change teachers during their school career. When lower-level units (i.e., an occasion) belong to different higher-level units (i.e., students and teachers) at the same time, we are confronted with cross-classified data. In cross-classified data, lower level units do not belong to one and only one higher-level unit. Rather, lower level units belong to pairs or combinations of higher-level units formed by crossing two or more higher-level classifications with one another. A special class of multilevel models, known as cross-classified random effects models (CCREM), has been developed for analysing data with such structure (Goldstein, 1987; Raudenbush, 1993).

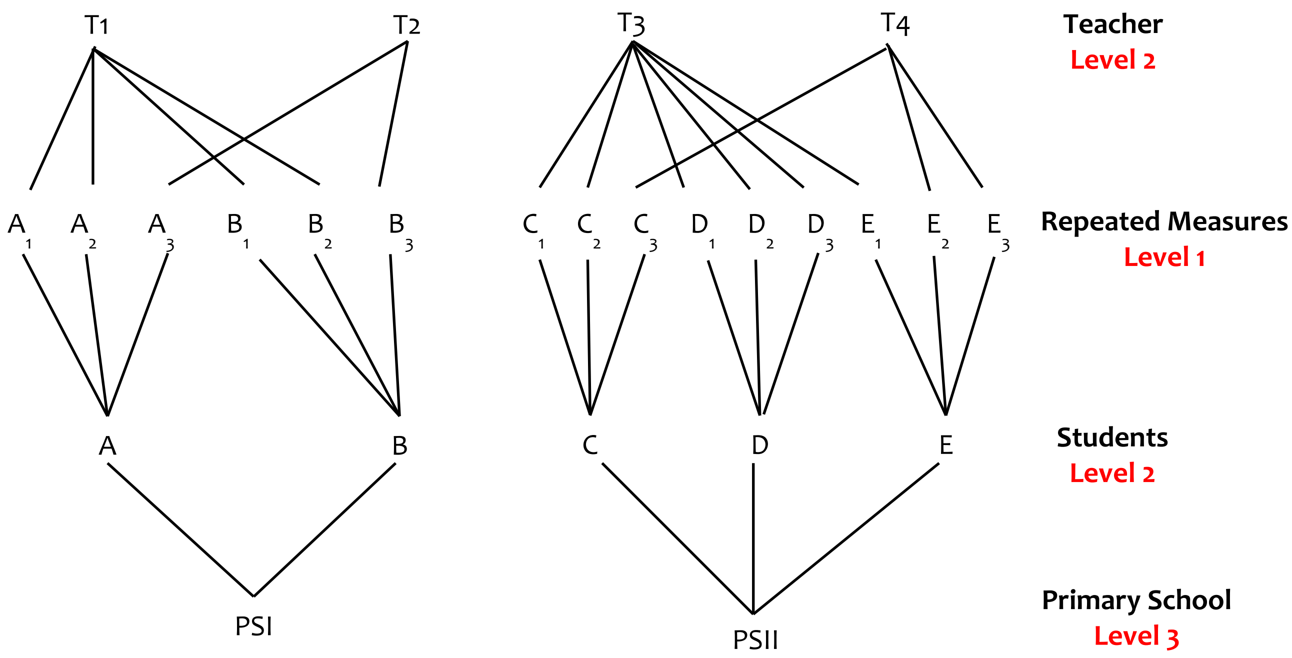
In this study, the model has to incorporate both students and teachers as sources of variation in achievement, but in such a manner that occasions are nested in the cross-classification of both students and teachers. The longitudinal data in our analytical sample are partially crossed because most students change teachers at least once during the three years under study (87.45% in language and 88.34% in mathematics) but not all students do so.

The following table shows how many students were taught by one, two and three teachers over the three years studied. When two teachers are associated to a particular student, three situations are possible; the same teacher taught the student the first two years and then another teacher does (AAB), there is a change of teacher after the first year and the second teacher continues over the third year (ABB), and, a situation that does not occur in these data, there is a change of teacher after the first year and then the first teacher taught the student again on the third year (ABA).

TABLE 19: SAMPLE SIZE BY NUMBER OF TEACHERS ASSOCIATED TO STUDENT

	LANGUAGE	MATHEMATICS
1 Teacher (AAA)	2,472 (12.55%)	2,297 (11.66%)
2 Teachers (AAB)	4,754 (24.13%)	4,828 (24.50%)
2 Teachers (ABB)	4,363 (22.14%)	4,570 (23.19%)
2 Teachers (ABA)	0 (0.00%)	0 (0.00%)
3 Teachers (ABC)	8,115 (41.18%)	8,009 (40.65%)
TOTAL	**19,704 (100.00%)**	**19,704 (100.00%)**

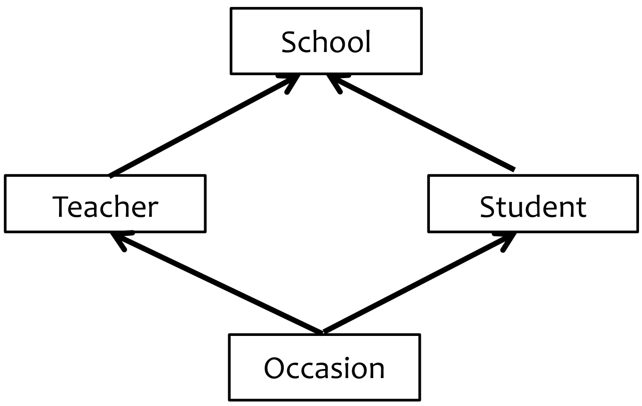
FIGURE 33: UNIT DIAGRAM FOR THE STUDENT-TEACHER CROSS-CLASSIFICATION



The unit diagram in Figure 33 conveys the two-way cross-classified structure of the data in terms of the actual units at each classification for a small number of cases. The solid lines show the exact relationships between every unit at every classification of the data structure. It is possible to observe that not all occasions from the same student are necessarily linked to the same teacher, nor do all occasions from the same teacher belong to the same student. Thus, the data have a two-way cross-classified non-hierarchical structure; the student and teacher hierarchies are crossed with one another. Also, there are two classifications in level 2 because students are not nested within teachers or teachers within students; the pure multilevel structure breaks down.

Clearly, for most data sets, unit diagrams can only be presented for small sub samples of the data as they quickly become unwieldy when many units are involved. Browne et al. (2001) proposed classification diagrams as an alternative to unit diagrams. Classification diagrams have the advantage of providing a simple summary of the data structure for the entire data set. Figure 34 presents the classification diagram for this study's two-way cross-classified data structure. These diagrams have one node for each classification in the model. Two nodes connected by a single arrow indicate a nested relationship, while two unconnected nodes indicate a crossed relationship.

FIGURE 34: CLASSIFICATION DIAGRAM FOR THE STUDENT-TEACHER CROSS-CLASSIFICATION



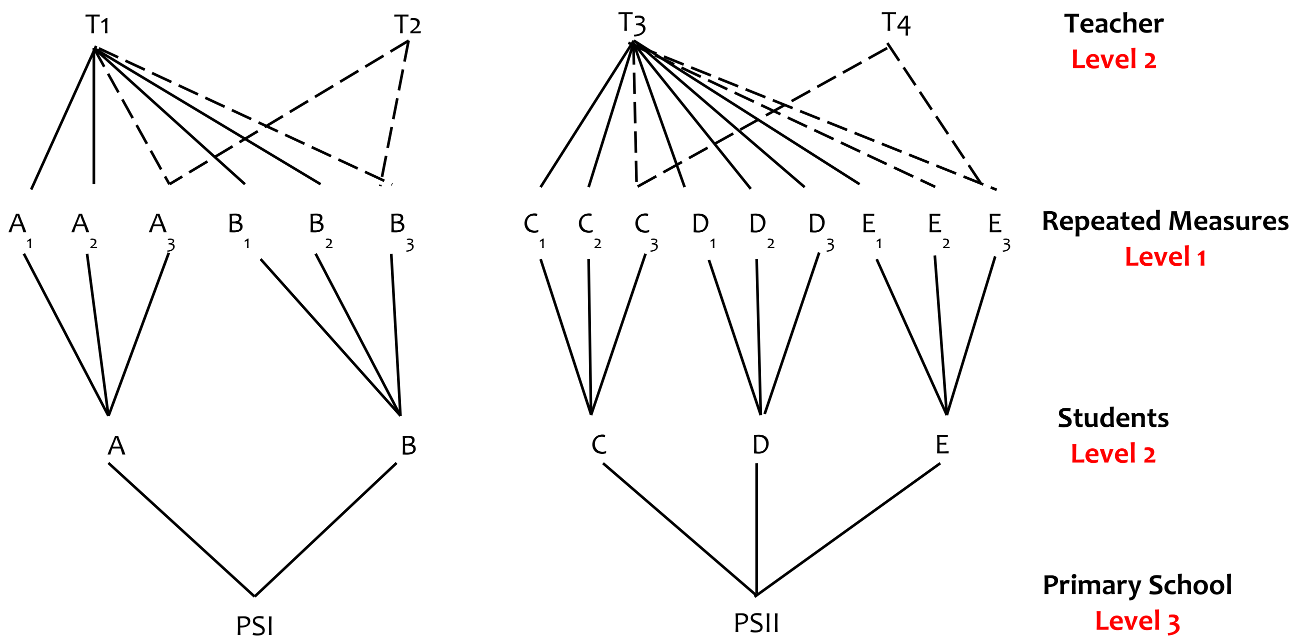
While this diagram again clearly shows the two-way cross-classified structure of the data, it is more abstract than the corresponding unit diagram (Figure 33) as it does not present the structure of the data in terms of the actual units at each classification. However, a considerable advantage of classification diagrams over unit diagrams is that they easily extend to more complex data structures that have many classifications.

6.3.3 *Multiple Membership Data*

As explained above, in longitudinal research, where students are assessed repeatedly and can change teachers during the period under study, data are cross-classified. Then, when confronted with student mobility, if students change teachers, then more than one teacher is related to their school success at a given time point. This situation represents another type of imperfect clustering that requires an extension of the conventional multilevel model. Since some students are members of multiple teachers, the data are said to have a multiple membership structure, this is, a situation where lower-level units (i.e., students' scores) belong to more than one higher-level unit of a population of interest (i.e., teachers).

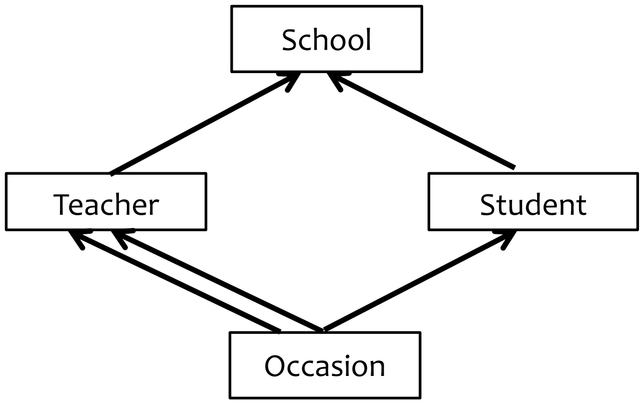
In these data, several pupils have been taught language and mathematics by more than one teacher during the period studied. Figure 35 presents this structure as a unit diagram. The unit diagram conveys the multiple membership structure of the data in terms of the actual units at each classification. The lines show the exact relationships between every unit at every classification of the data structure. Solid lines denote nested relationships (i.e., scores who belong to one teacher), while dashed lines denote multiple membership relationships (i.e., scores belonging to two or more teachers as these are the scores of students after having changed teachers).

FIGURE 35: UNIT DIAGRAM FOR THE MULTIPLE MEMBERSHIP OF STUDENTS' MEASUREMENT OCCASIONS TO TEACHERS



The classification diagram for this situation is depicted in Figure 36 where a double arrow is used to indicate the multiple membership structure.

FIGURE 36: CLASSIFICATION DIAGRAM FOR THE MULTIPLE MEMBERSHIP OF STUDENTS' MEASUREMENT OCCASIONS TO TEACHERS



An important feature of multiple membership data structures is that the degree to which each lower level unit belongs to each higher-level unit will often vary across those higher-level units, in this case teachers. Multiple membership weights are used to quantify this and the information is used when fitting multiple membership models. In this study students may spend more time with some teachers than others. Here, multiple membership weights are defined as the proportion of time spent with each teacher. Thus, if a student is taught for one year by teacher A and then for two years by teacher B, multiple membership weights would be assigned proportionally: 0.33 and 0.66 for teachers A and B respectively. These weights reflect the fact that we might expect teacher B to be more influential in determining the student's outcome than teacher A.

Researchers typically use one of two procedures for handling complex multilevel data structures such as cross-classified and multiple membership data (Beretvas, 2011). As a first strategy, researchers might delete from analysis the sets of units that prevent the data from being a pure hierarchy (this is, deleting mobile students from the datasets being analysed). As an alternative strategy, researchers could ignore one of the cross-classified factors or all but one higher-level unit associated with multiple member units.

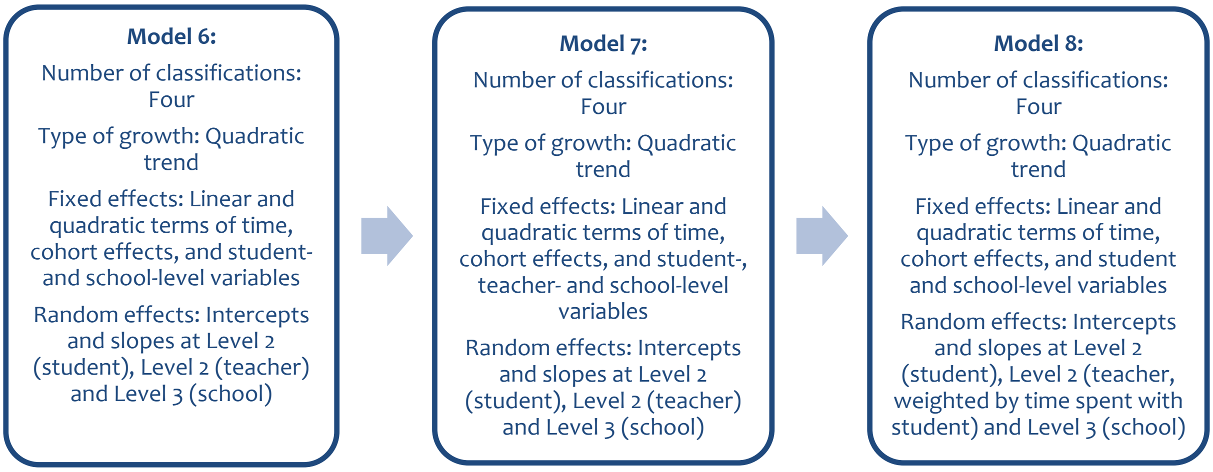
However, deleting cases reduces power unnecessarily and can affect generalisability and validity of inferences (Meyers and Beretvas, 2006). Also, ignoring one of the classification factors or all but one higher-level unit associated with multiple member units can lead to inaccurate variance component estimation (Fielding and Goldstein, 2006; Rasbash and Browne, 2001). Indeed, naively fitting the nearest equivalent hierarchical model to cross-classified data will lead us to misattribute response variation to the included levels (Moerbeek, 2004; van den Noortgate, Opdenakker and Onghena, 2005; van Landeghem, De Fraine and van Damme, 2005). This in turn may lead to drawing misleading conclusions about the relative importance of different sources of influence on the response. Similarly, assigning students to the first teacher that teaches them and then fitting a students-within-teachers model of student attainment will likely lead to underestimate the importance of teachers and overestimate the importance of students as sources of variation in student attainment.

Thus, deleting cases and ignoring complex structures removes the problem and handle the remaining dependency, but compromises the validity of inferences. Instead, it is recommended that the cross-classified and multiple membership nature of the data are correctly modelled, when present. To do this, researchers could use one of two extensions to the conventional multilevel model: the Cross-Classified Random Effects Model (CCREM) and the Cross-Classified Multiple Membership Random Effects Model (CCMMREM). These models are applied in the following sections.

6.3.4 Models

Figure 37 shows the progression of models fitted in this chapter. The analyses are organised in three models. Model 6 is a two-way cross-classified model in which student and teacher hierarchies are crossed with one another and nested within schools. Model 7 includes the effect of teacher predictors (i.e., gender, initial teacher training duration, major and years of teaching experience) on student achievement status and growth. Finally, Model 8 is a cross-classified multiple membership model that assumes that the effects of teachers from previous years are carried forward to the following years.

FIGURE 37: GROW TH MODELS IN CHAPTER 6



6.3.4.1 *Model 6: A cross-classified accelerated growth model with a quadratic effect of time, and student- and school-level covariates*

Crossed-random effects are added to the model specifications in the previous chapter to investigate the magnitude of teacher effects. The following equation denotes this model (Model 6), accompanied with the parameters brought forward from Model 4 in Chapter 5. Multiple subscript notation, introduced by Rasbash and Browne (2001), is adopted in the formulation of this model as it facilitates the description of multilevel models with combinations of hierarchical, crossed and multiple membership structures.

EQUATION 10

$$Y_{t(i_1,i_2)j} = \beta_{0(i_1,i_2)j} + \beta_{1(i_1,i_2)j} t_{t(i_1,i_2)j} + \beta_2 t^2_{t(i_1,i_2)j} + e_{t(i_1,i_2)j}$$

$$\begin{aligned}\beta_{0(i_1,i_2)j} = \beta_{000j} &+ \beta_{0010} Cohort2_{i_1j} + \beta_{0020} Cohort3_{i_1j} + \beta_{0030} Cohort4_{i_1j} \\ &+ \beta_{0040} Female_{i_1j} + \beta_{0050} Age_{i_1j} + \beta_{0060} SES_{i_1j} + \beta_{0070} Books_{i_1j} + u_{0i_1j} \\ &+ u_{00i_2j}\end{aligned}$$

$$\begin{aligned}\beta_{1(i_1,i_2)j} = \beta_{010j} & \\ &+ \beta_{0110} Cohort2_{i_1j} + \beta_{0120} Cohort3_{i_1j} + \beta_{0130} Cohort4_{i_1j} \\ &+ \beta_{0140} Female_{i_1j} + \beta_{0150} Age_{i_1j} + \beta_{0160} SES_{i_1j} + \beta_{0170} Books_{i_1j} + u_{1i_1j} \\ &+ u_{10i_2j}\end{aligned}$$

$$\begin{aligned}\beta_{000j} = \beta_{0000} &+ \beta_{0100} Achievement\ Mean_j + \beta_{0200} Achievement\ SD_j \\ &+ \beta_{0300} School\ SES_j + u_{000j}\end{aligned}$$

$$\begin{aligned}\beta_{010j} = \beta_{1000} &+ \beta_{1100} Achievement\ Mean_j + \beta_{1200} Achievement\ SD_j \\ &+ \beta_{1300} School\ SES_j + u_{100j}\end{aligned}$$

$$\begin{pmatrix} u_{0i_1j} \\ u_{1i_1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right]$$
$$\begin{pmatrix} u_{00i_2j} \\ u_{10i_2j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u00}^2 & \\ \sigma_{u00u10} & \sigma_{u10}^2 \end{pmatrix}\right]$$
$$\begin{pmatrix} u_{000j} \\ u_{100j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u000}^2 & \\ \sigma_{u000u100} & \sigma_{u100}^2 \end{pmatrix}\right]$$
$$e_{t(i_1,i_2)j} \sim N(0, \sigma_e^2).$$

The number of letters in the subscript identifies the number of classifications (here, there are four: occasion, student, teacher, and school). Subscripts with the same common letter (here, i_1 and i_2) appearing in parenthesis separated by a comma identify cross-classified factors at the same level. The student, i_1, appears before the teacher identifier, i_2. $Y_{t(i_1,i_2)j}$ represents the score at time t of student i_1, with teacher i_2, from school j. Each residual is assumed normally distributed with a mean of zero and its own variance. Covariances among residuals in different classifications are assumed to be zero.

Independent variables introduced in Model 4 (i.e., *Female*, *Age*, *SES*, *Number of books at home*, *Achievement Mean, Achievement SD* and *School SES*) are retained in the models fitted in this chapter as both fixed effects and in interaction with the *Time* variable.

6.3.4.2 *Model 7: A cross-classified accelerated growth model with a quadratic effect of time, and student-, teacher- and school-level covariates*

As shown in Equation 11, teacher-level predictors are included in Model 7 to depict the effects of teacher characteristics on student achievement status and growth. These variables are *Female Teacher*, which refers to teacher's gender, and the baseline category is "male", *ITT Duration*, indicating the duration of the teacher's initial teacher training programme in semesters, *Experience*, denoting the number of years that the educator has been teaching in any school and, finally, *Major*, referring to whether the teacher has undertaken specialised training in a particular subject or not. When modelling student progress in language, this variable refers to whether the training of the teacher associated to the student included a major in language. Correspondingly, if modelling student progress in mathematics, this variable indicates whether the teacher was trained in a programme including a major in mathematics. The baseline of this dichotomous variable is the condition without major. These variables were accessed via the Teacher Census data sets. Appendix 4 shows the operational definition and descriptive statistics for these teacher-level variables. Continuous variables were grand-mean centered.

EQUATION 11

$$Y_{t(i_1,i_2)j} = \beta_{0(i_1,i_2)j} + \beta_{1(i_1,i_2)j}t_{t(i_1,i_2)j} + \beta_2 t^2_{t(i_1,i_2)j} + e_{t(i_1,i_2)j}$$

$$\begin{aligned}\beta_{0(i_1,i_2)j} = {} & \beta_{000j} + \beta_{0010}Cohort2_{i_1j} + \beta_{0020}Cohort3_{i_1j} + \beta_{0030}Cohort4_{i_1j} \\ & + \beta_{0040}Female_{i_1j} + \beta_{0050}Age_{i_1j} + \beta_{0060}SES_{i_1j} + \beta_{0070}Books_{i_1j} \\ & + \beta_{0080}Female\ Teacher_{i_2j} + \beta_{0090}ITT\ Duration_{i_2j} + \beta_{0100}Major_{i_2j} \\ & + \beta_{0110}Experience_{i_2j} + u_{0i_1j} + u_{00i_2j}\end{aligned}$$

$$\begin{aligned}\beta_{1(i_1,i_2)j} = {} & \beta_{010j} \\ & + \beta_{0120}Cohort2_{i_1j} + \beta_{0130}Cohort3_{i_1j} + \beta_{0140}Cohort4_{i_1j} \\ & + \beta_{0150}Female_{i_1j} + \beta_{0160}Age_{i_1j} + \beta_{0170}SES_{i_1j} + \beta_{0180}Books_{i_1j} \\ & + \beta_{0190}Female\ Teacher_{i_2j} + \beta_{0200}ITT\ Duration_{i_2j} + \beta_{0210}Major_{i_2j} \\ & + \beta_{0220}Experience_{i_2j} + u_{1i_1j} + u_{10i_2j}\end{aligned}$$

$$\begin{aligned}\beta_{000j} = {} & \beta_{0000} + \beta_{1000}Achievement\ Mean_j + \beta_{1100}Achievement\ SD_j \\ & + \beta_{1200}School\ SES_j + u_{000j}\end{aligned}$$

$$\begin{aligned}\beta_{010j} = {} & \beta_{1300} + \beta_{1400}Achievement\ Mean_j + \beta_{1500}Achievement\ SD_j \\ & + \beta_{1600}School\ SES_j + u_{100j}\end{aligned}$$

$$\begin{pmatrix} u_{0i_1j} \\ u_{1i_1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & \\ \sigma_{u0u1} & \sigma^2_{u1} \end{pmatrix}\right]$$

$$\begin{pmatrix} u_{00i_2j} \\ u_{10i_2j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u00} & \\ \sigma_{u00u10} & \sigma^2_{u10} \end{pmatrix}\right]$$

$$\begin{pmatrix} u_{000j} \\ u_{100j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u000} & \\ \sigma_{u000u100} & \sigma^2_{u100} \end{pmatrix}\right]$$

$$e_{t(i_1,i_2)j} \sim N(0, \sigma^2_e).$$

6.3.4.3 *MODEL 8: A CROSS-CLASSIFIED MULTIPLE MEMBERSHIP ACCELERATED GROWTH MODEL WITH A QUADRATIC EFFECT OF TIME, AND STUDENT- AND SCHOOL-LEVEL COVARIATES*

Cross-classified and multiple membership data structures can both be present and of interest to the researcher. In this case, repeated measures are cross-classified on students and teachers, and nested by primary schools attended. Furthermore, since some students change teachers over the period under study, their score at a given time point belongs to more than one teacher. Cross-Classified Multiple Membership Random Effects Model (CCMMREM) allows us to depict both situations at the same time

(Beretvas, 2011). By adding the specifications from our previous models to CCMMREM we obtain the accelerated cross-classified multiple membership growth model shown in Equation 12, referred to as Model 8.

EQUATION 12

$$Y_{t(i_1,\{i_2\})j} = \beta_{0(i_1,\{i_2\})j} + \beta_{1(i_1,\{i_2\})j} t_{t(i_1,\{i_2\})j} + \beta_2 t^2_{t(i_1,\{i_2\})j} + e_{t(i_1,\{i_2\})j}$$

$$\beta_{0(i_1,\{i_2\})j} = \beta_{000j} + \beta_{0010} Cohort2_{i_1j} + \beta_{0020} Cohort3_{i_1j} + \beta_{0030} Cohort4_{i_1j} + \beta_{0040} Female_{i_1j} + \beta_{0050} SES_{i_1j} + \beta_{0060} Books_{0i_1j} + \beta_{0070} Age_{0i_1j} + u_{0i_1j} + \sum_{h\in\{i_2\}} w_{thj} u_{00hj}$$

$$\beta_{1(i_1,\{i_2\})j} = \beta_{010j} + \beta_{0110} Cohort2_{i_1j} + \beta_{0120} Cohort3_{i_1j} + \beta_{0130} Cohort4_{i_1j} + \beta_{0140} Female_{i_1j} + \beta_{0150} SES_{i_1j} + \beta_{0160} Books_{i_1j} + \beta_{0170} Age_{i_1j} + u_{1i_1j} + \sum_{h\in\{i_2\}} w_{thj} u_{10hj}$$

$$\beta_{000j} = \beta_{0000} + \beta_{0100} Achievement\ Mean_j + \beta_{0200} Achievement\ SD_j + u_{000j}$$

$$\beta_{010j} = \beta_{1000} + \beta_{1100} Achievement\ Mean_j + \beta_{1200} Achievement\ SD_j + u_{100j}$$

$$\begin{pmatrix} u_{0i_1j} \\ u_{1i_1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & \\ \sigma_{u0u1} & \sigma^2_{u1} \end{pmatrix}\right]$$

$$\begin{pmatrix} u_{00i_2j} \\ u_{10i_2j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u00} & \\ \sigma_{u00u10} & \sigma^2_{u10} \end{pmatrix}\right]$$

$$\begin{pmatrix} u_{000j} \\ u_{100j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u000} & \\ \sigma_{u000u100} & \sigma^2_{u100} \end{pmatrix}\right]$$

$$e_{t(i_1,i_2)j} \sim N(0, \sigma^2_e).$$

Here, $Y_{t(i_1,\{i_2\})j}$ represents the score at time t of student $i_1$that was taught by a set of i_2 teachers in school j. Thus, t indexes the level-1 unit (occasions) that is a member of multiple units of the level-2 classification i_2 (teachers). u_{00hj} and u_{10hj} are the residuals associated with teachers and w_{thj} is the weight assigned to occasions' association with teachers.

Multiple membership data are modelled using weighting, where the membership weights are usually proportional to the time a lower-level unit spent at a higher-level unit, with the weights summing to 1. In this study, a pupil who was taught by the same teacher from year level 3 to Year level 5 has a membership weight of 1 for that teacher and 0 for all other teachers. A pupil taught by a different teacher each of the three in which data was collected has a membership weight of 1/3 for each of them and 0 for all other teachers[39].

The fit of Model 6, which assumes that teacher contributions disappear at the end of each year, is compared to that of Model 8 that assumes that the effects of teachers from previous years are entirely carried forward to the following years. By doing this it is possible to test whether there are indications of cumulative effects of teachers in these data.

6.4 Results

This chapter addresses general issues in the analysis of teacher effects in Chile, calculated as the variation in teacher value-added scores. The main aspects investigated are: the size of teacher effects on students' achievement growth in language and mathematics, the consistency of teacher effects across different academic subjects (language and mathematics), the teacher-level predictors of these effects and to what extent they accumulate over time.

As explained in the previous section, the complex structure of the data is acknowledged using a cross-classified model, which addresses the fact that repeated measurements of achievement test scores are nested within both students and teachers (Model 6) and a cross-classified multiple membership model, in order to test the hypothesis of teachers' cumulative effects (Model 8). Thus, the models combine these aspects with the accelerated growth model introduced in Chapter 4, obtaining a cross-classified accelerated growth model and a cross-classified multiple membership accelerated growth model. Results derived from these models are presented in this section.

[39] Other weighting schemes may also be applied, and it might be interesting to explore a range of weighting schemes as part of a sensitivity analysis for the model. For example, it is possible to think that the order in which teachers have taught is important. In particular we might want to allow the final teacher attended to have a greater influence than earlier teachers. Alternatively, we might reason that each year being taught by a teacher during primary schooling is more important than the last. However, these options are not tested in this study.

6.4.1 *The Magnitude of Teacher Effects*

Table 20 shows the results of Model 6: the cross-classified accelerated growth model. With regard to the fixed part of the model, the effects of the student and school variables that were introduced in Model 4 (in Chapter 5) persist in Model 6. The DIC values indicate that there is a sizeable improvement in the model fit when the teacher classification is incorporated and the cross-classified structure of the data acknowledged, as shown by the reduction on DIC values from Model 4 (three-level accelerated growth model without the student-teacher cross-classification) to Model 6 (ΔDIC = -3,495 in language, and ΔDIC = -3,051 in mathematics).

In addition, the variance components on achievement growth rates show that most of the variance in growth appears to be at the teacher level (53.5% and 66.2% for the linear component in language and mathematics, respectively). The variance at the school level is notably smaller that the variance at the teacher level (21.0% and 29.8% in language and mathematics, respectively), which has been suggested in previous research (Hattie, 2009; Muijs et al., 2014; Scheerens and Bosker, 1997), although it is still substantial and larger than estimates in non growth curve models across two time points found in the literature. Furthermore, the teacher effects yield a large *d*-type effect size of 0.731 and 0.814 for achievement growth in language and mathematics, respectively[40].

In both language and maths, teacher effects on the linear growth rate are larger than on students' achievement status (intercept) (which are 8.6% and 11.0%, respectively, figures in line with past research). In addition, teacher effects on achievement growth are larger for mathematics than for language. This result is in line with previous research suggesting that school and teacher effects are larger in subject areas that are typically learned at school, as with mathematics, where exposure is limited in the family and the community (Teddlie and Reynolds, 2000; Thomas et al., 1997b). Furthermore, these educational effects are similar and even larger than those found in previous studies using alike model specifications for similar outcomes (i.e., Palardy, 2010; Rowan, Correnti and Miller, 2002).

---

[40] The *d*-type effect size is calculated as proposed by Rowan et al. (2002). Thus, for Model 5:

$$d = \frac{\sqrt{\sigma^2_{u_{10i_2j}}}}{\sqrt{(\sigma^2_{u_{100j}} + \sigma^2_{u_{10i_2j}} + \sigma^2_{u_{1i_1j}})}}.$$

TABLE 20: FULL MODEL 6 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 6 LANGUAGE		MODEL 6 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Fixed Part				
Intercept	348.363 (0.603)	347.154 - 349.572	344.977 (0.644)	343.665 - 346.290
Time	34.295 (1.116)	31.718 - 36.872	31.522 (1.257)	28.551 - 34.494
Time2	-7.574 (0.592)	-9.004 - -6.143	-8.132 (0.758)	-10.039 - -6.225
Cohort 2	25.305 (0.565)	24.187 - 26.423	23.042 (0.533)	21.985 - 24.100
Cohort 3	37.217 (0.616)	35.994 - 38.439	34.432 (0.604)	33.228 - 35.636
Cohort 4	55.187 (0.616)	53.964 - 56.409	56.923 (0.634)	55.645 - 58.201
Time x Cohort 2	-10.399 (1.269)	-13.160 - -7.637	-15.576 (1.264)	-18.325 - -12.826
Time x Cohort 3	-10.121 (1.195)	-12.694 - -7.548	-5.806 (1.881)	-10.403 - -1.208
Time x Cohort 4	-14.003 (1.202)	-16.580 - -11.426	-4.622 (1.894)	-9.252 - 0.008
Time2 x Cohort 2	3.039 (0.57)	1.797 - 4.281	8.058 (0.612)	6.700 - 9.416
Time2 x Cohort 3	2.701 (0.582)	1.427 - 3.975	5.518 (0.941)	3.189 - 7.846
Time2 x Cohort 4	6.312 (0.585)	5.030 - 7.593	6.127 (0.911)	3.882 - 8.373
Female student	1.949 (0.313)	1.317 - 2.580	-2.103 (0.300)	-2.714 - -1.491
Age	-1.689 (0.267)	-2.221 - -1.158	-1.665 (0.273)	-2.223 - -1.107
SES	0.761 (0.300)	0.070 - 1.451	0.870 (0.226)	0.385 - 1.354
Number books at home	2.337 (0.150)	2.043 - 2.631	2.023 (0.171)	1.672 - 2.374
Time x Female	0.376 (0.203)	-0.079 - 0.832	-0.663 (0.146)	-0.963 - -0.362
Time x Age	-0.351 (0.129)	-0.610 - -0.093	-0.462 (0.208)	-0.955 - 0.032
Time x SES	-0.063 (0.278)	-0.751 - 0.626	-0.121 (0.139)	-0.408 - 0.166
Time x Number books at home	-0.205 (0.107)	-0.443 - 0.034	-0.095 (0.086)	-0.278 - 0.087
Achievement Mean	0.218 (0.035)	0.148 - 0.289	0.275 (0.030)	0.214 - 0.336
Achievement SD	-0.204 (0.131)	-0.463 - 0.054	-0.143 (0.106)	-0.352 - 0.066
School SES	2.403 (0.821)	0.692 - 4.114	0.419 (0.840)	-1.317 - 2.155
Time x Achievement Mean	-0.052 (0.023)	-0.099 - -0.005	-0.037 (0.028)	-0.098 - 0.025
Time x Achievement SD	-0.146 (0.113)	-0.393 - 0.101	-0.082 (0.079)	-0.242 - 0.078
Time x School SES	-0.176 (0.440)	-1.059 - 0.706	0.700 (0.661)	-0.763 - 2.163
Random Part				

	MODEL 6 LANGUAGE		MODEL 6 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Random Level 4 (School)				
Intercept	15.499 (3.425)	8.532 - 22.465	15.642 (4.451)	5.898 - 25.385
Time/Intercept	-3.141 (1.499)	-6.103 - -0.178	-4.064 (1.802)	-7.677 - -0.451
Time	4.653 (1.687)	0.995 - 8.312	5.418 (1.532)	2.326 - 8.510
Random Level 3 (Teacher)				
Intercept	20.837 (3.053)	14.125 - 27.549	22.218 (3.014)	15.751 - 28.684
Time/Intercept	-12.320 (2.021)	-16.614 - -8.025	-11.832 (2.118)	-16.343 - -7.322
Time	11.868 (1.987)	7.514 - 16.222	12.059 (2.300)	6.901 - 17.216
Random Level 2 (Student)				
Intercept	204.562 (3.307)	198.072 - 211.052	163.654 (3.325)	156.795 - 170.514
Time/Intercept	-11.400 (1.582)	-14.659 - -8.140	-2.188 (0.980)	-4.181 - -0.196
Time	5.675 (1.075)	3.449 - 7.900	0.732 (0.240)	0.261 - 1.203
Random Level 1 (Occasion)				
Intercept	109.269 (1.273)	106.734 - 111.805	105.535 (0.901)	103.744 - 107.326
School Initial Status – Growth Correlation	-0.370		-0.441	
Teacher Initial Status – Growth Correlation	-0.783		-0.723	
Student Initial Status – Growth Correlation	-0.335		-0.200	
School ICC Initial Status	0.064		0.078	
School ICC Growth	0.210		0.298	
Teacher ICC Initial Status	0.086		0.110	
Teacher ICC Growth	0.535		0.662	
DIC	464,396		460,157	
pD	18,977		17,171	
Units: Schools	**156**		**156**	
Units: Teachers	**851**		**812**	
Units: Students	**19,704**		**19,704**	
Units: Occasions	**59,112**		**59,112**	

FIGURE 38: TEACHER LINES BASED ON MODEL 6

LANGUAGE MATHEMATICS

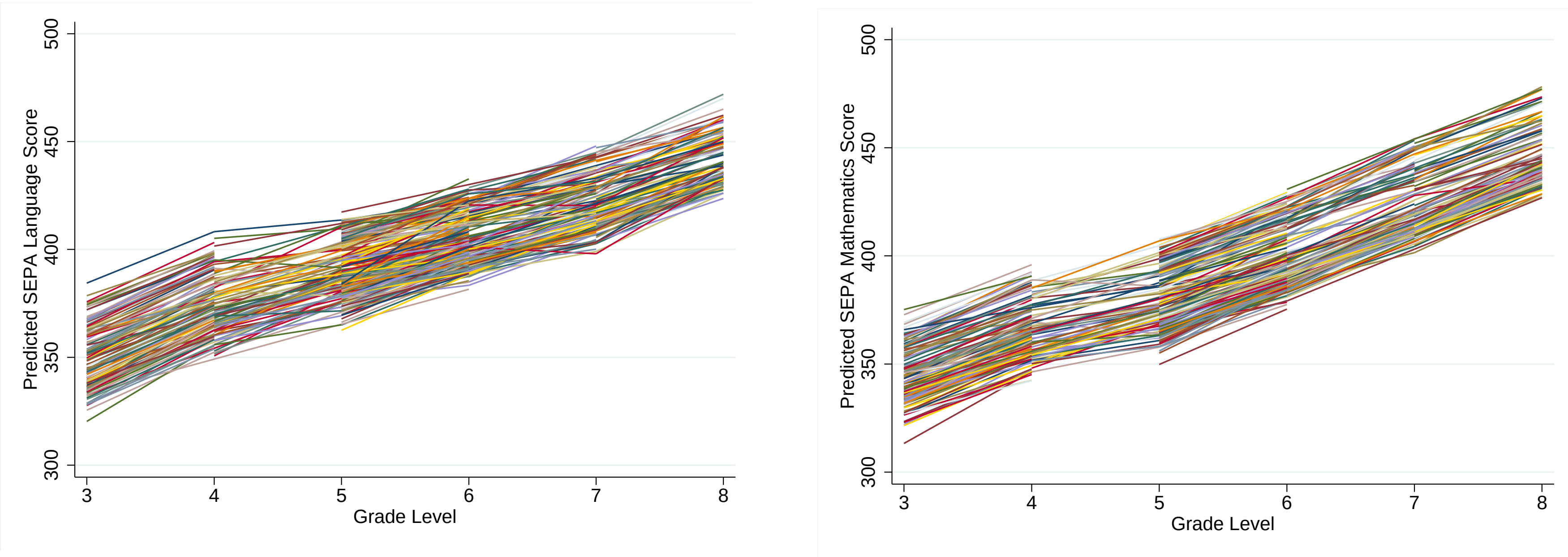
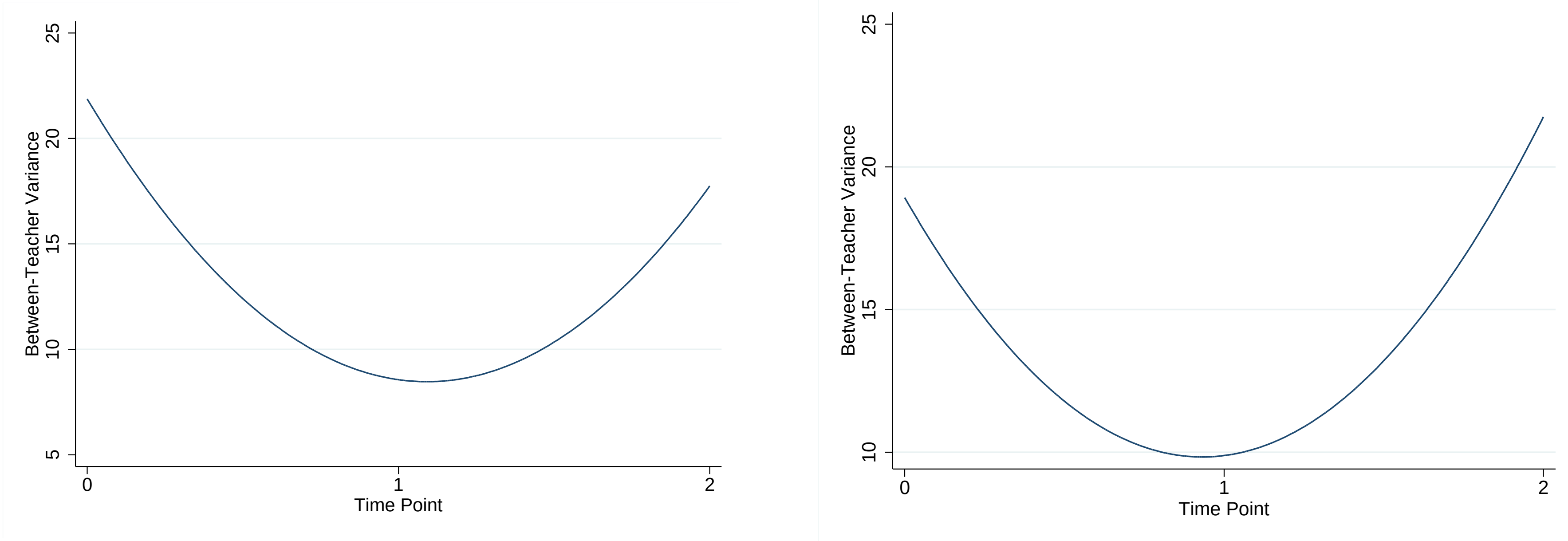


FIGURE 39: TEACHER-LEVEL VARIANCE FUNCTION

LANGUAGE MATHEMATICS



To illustrate the magnitude of teacher effects on achievement status and growth, the 851 teacher residuals in the language sample and the 812 residuals for teachers in the mathematics sample are plotted in Figures 40 and 41. These caterpillar plots graph each residual, obtained from Model 6, against their rank order, accompanied by error bars corresponding to confidence intervals (Goldstein and Healy, 1995).

In both subjects there is considerable overlap of intervals, so that only widely separated teachers can be judged as having significantly different effects on students' achievement growth. Generally speaking, teacher effects are not estimated with great precision. Nonetheless, it is possible to distinguish some outliers at each end. In language, the confidence intervals of the residuals do not overlap zero for a group of about 46 teachers at the lower end of the intercept residual plot and for 46 at the upper end. In addition, about 25 and 32 teachers are at the lower and upper end of the language slope residual plot. In mathematics, the numbers of outlier teachers are 37 and 47 at the lower and upper end of the intercept residual plot, respectively. In the mathematics slope residual plots, in turn, there are 18 and 17 at the lower and upper end. This means that approximately 4 to 11% of the teachers differ significantly from the average teacher effect at the 5% level.

In Table 21 the 572 teachers in the sample that taught both language and mathematics have been classified by categorising their residuals for the two outcome measures. If the confidence interval for a teacher residual does not overlap zero, the value added residual is significantly different either above or below expectation, and the teacher is identified as an outlier. The table suggests that teachers that are effective in one subject are not necessarily effective in the other subject. The consistency of teacher effects is analysed in greater detail in the following section.

Finally, the estimated residuals specified in Model 6 were used to check an important model assumption; that the residuals at each level follow normal distributions. To this end, normal probability plots, in which the ranked residuals are plotted against corresponding points on a normal distribution curve (see figures in Appendix 7). The plots look fairly linear, which suggests that the assumption of normality is reasonable.

FIGURE 40: TEACHER RESIDUALS FOR LANGUAGE BASED ON MODEL 6

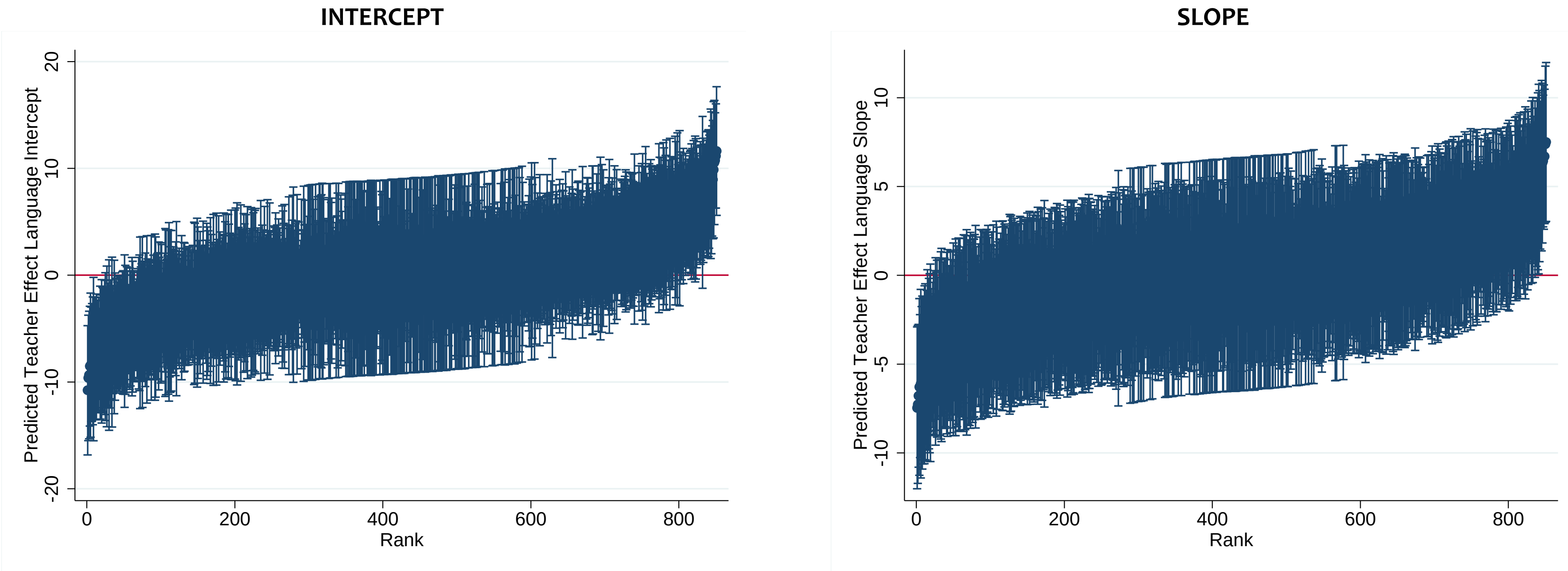


FIGURE 41: TEACHER RESIDUALS FOR MATHEMATICS BASED ON MODEL 6

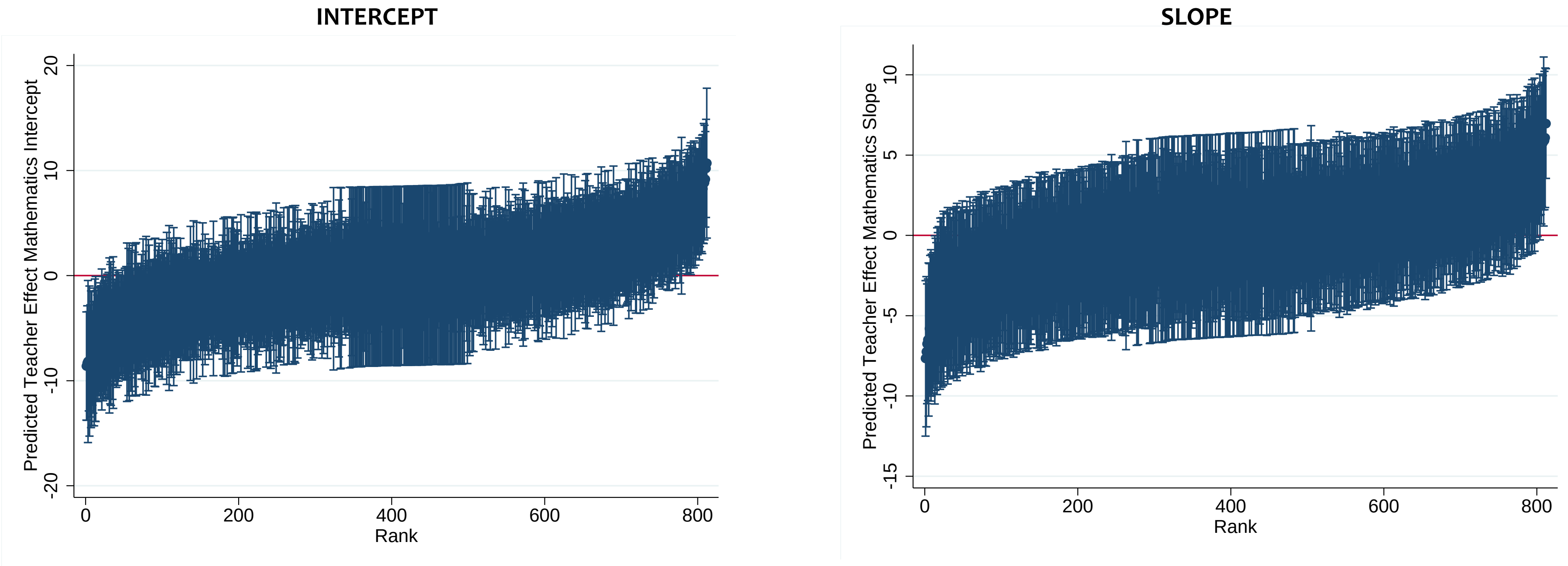


TABLE 21: CROSS TABULATION OF TEACHER EFFECTS ON STUDENT ACHIEVEMENT GROWTH

	LANGUAGE			
MATHEMATICS	Significant ++	Non-Significant +	Non-Significant -	Significant - -
Significant ++	0 (0.00%)	1 (0.17%)	2 (0.35%)	0 (0.00%)
Non-Significant +	5 (0.87%)	151 (26.40%)	101 (17.66%)	2 (0.35%)
Non-Significant -	4 (0.70%)	105 (18.36%)	179 (31.29%)	10 (1.75%)
Significant - -	0 (0.00%)	4 (0.70%)	8 (1.40%)	0 (0.00%)

Note: Percentages given in brackets.

6.4.2 *Consistency of Teacher Effects across Subjects*

Given the sparseness of prior research on the topic of consistency of teacher effects across subjects, this study's data provide relevant insights. To assess whether teachers had consistent effects on students' achievement across different academic subjects, the correlations among teacher residuals from Model 6 were calculated.

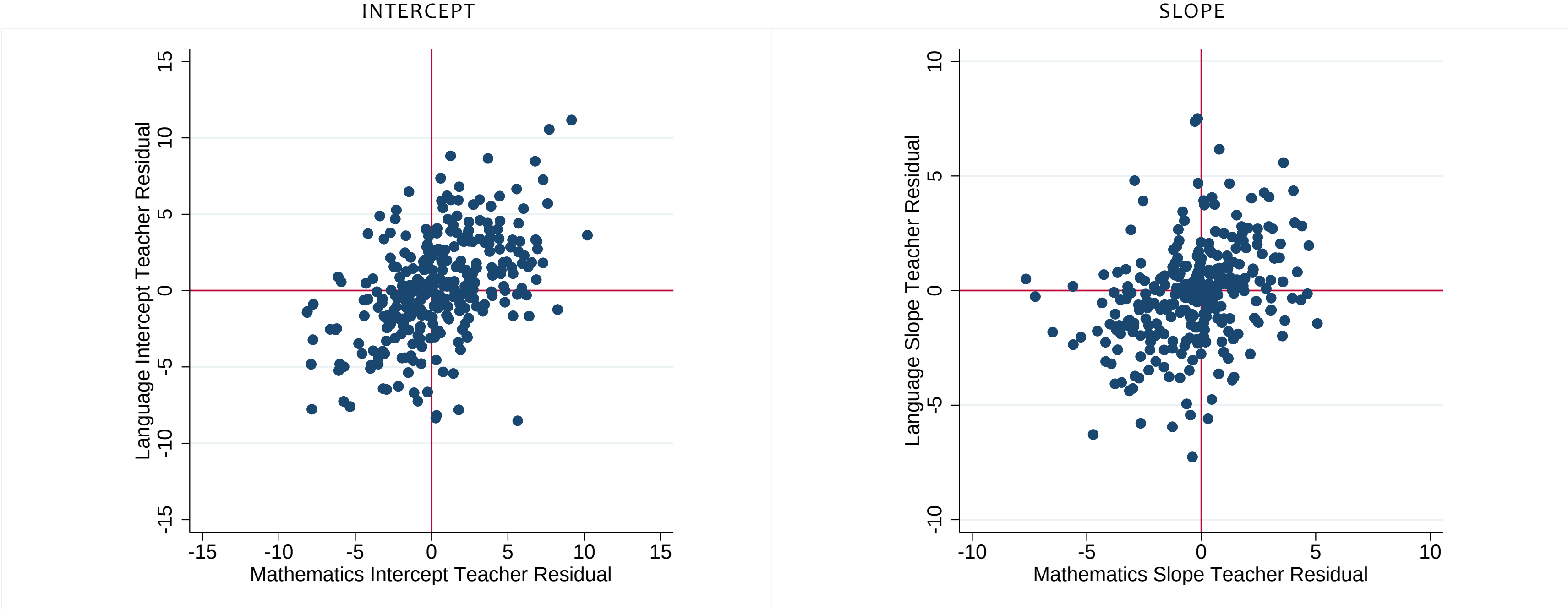
TABLE 22: CORRELATION BETWEEN MODEL 6 TEACHER RESIDUALS

	Language Intercept Residual	Language Slope Residual	Mathematics Intercept Residual	Mathematics Slope Residual
Language Intercept Residual	1			
Language Slope Residual	-0.874***	1		
Mathematics Intercept Residual	0.481***	-0.404***	1	
Mathematics Slope Residual	-0.360***	0.346***	-0.793***	1

Note: Pearson correlation was used. Significance of parameter estimates is indicated by asterisks: *** = p<.001. $N_{teachers}$ = 572.

As shown in Table 22, for the 572 teachers in our sample that taught both language and mathematics there is a moderate degree of consistency in the effects across language and mathematics achievement (*r*=0.481 for intercept residuals and *r*=0.346 for slope residuals). The results therefore suggest that a given teacher varies in effectiveness when teaching different academic subjects. Scatterplots in Figure 42 also depicts the relationship of teacher residuals across subjects. The following section will consider the role of teacher input predictors in explaining the large variance found at the teacher level.

FIGURE 42: SCATTERPLOTS OF LANGUAGE AND MATHEMATICS TEACHER RESIDUALS BASED ON MODEL 6



6.4.3 *TEACHER-LEVEL PREDICTORS*

To this point, the focus has been on evidence on the overall size of teacher effects on student achievement. These estimates, although informative about how the Chilean educational system works, do not provide any evidence about why some teachers are more instructionally effective than others. To explain this phenomenon, it is necessary to inquire about the properties of teachers and their teaching that produce effects on students' growth in achievement. While the data available does not allow the exploration of teaching effects, namely the specific classroom processes that might explain these large effects, this section focuses on examining the effect of teacher-level characteristics that might explain part of the large variation found across teachers.

In Model 7, the teacher variables of interest are *Female Teacher, ITT Duration, Major* and *Experience*. Including these variables does not lead to an improvement in model fit, as the DIC value increases, for both subjects, in comparison to Model 6 (ΔDIC = 12 in language, and ΔDIC = 21 in mathematics). The addition of the teacher-level variables explained only 5% of the teacher-level variance in achievement status observed in Model 6, for both subjects, and a negligible proportion of the teacher-level variance on achievement growth.

As shown in Table 23, teacher variables that were found to be associated with achievement status were *Female Teacher* and *Major*, although only in mathematics, suggesting that students with female teachers and with teachers with no major in mathematics present higher achievement levels in mathematics in Grade 3. Also, only *Major* was found to be associated with achievement growth, indicating that students with teachers that hold a major in mathematics show higher growth rates in the subject. This suggests that subject expertise in mathematics is important and associated with higher student achievement in latter stages of primary school. In Figures 43 to 46, these trends are depicted, showing the differences in the relative positions of students with teachers presenting different characteristics.

TABLE 23: FULL MODEL 7 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 7 LANGUAGE		MODEL 7 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Fixed Part				
Intercept	347.128 (0.910)	345.285 - 348.971	344.053 (0.949)	342.029 - 346.076
Time	34.914 (1.418)	31.601 - 38.227	31.864 (1.464)	28.399 - 35.329
Time2	-7.549 (0.615)	-9.051 - -6.048	-8.197 (0.785)	-10.179 - -6.215
Cohort 2	25.302 (0.571)	24.170 - 26.434	23.075 (0.534)	22.016 - 24.135
Cohort 3	37.350 (0.647)	36.052 - 38.648	34.914 (0.608)	33.705 - 36.123
Cohort 4	55.343 (0.648)	54.042 - 56.644	57.422 (0.641)	56.129 - 58.715
Time x Cohort 2	-10.247 (1.250)	-12.954 - -7.539	-15.333 (1.270)	-18.089 - -12.577
Time x Cohort 3	-10.099 (1.240)	-12.800 - -7.398	-6.182 (1.897)	-10.810 - -1.554
Time x Cohort 4	-14.031 (1.217)	-16.650 - -11.413	-4.983 (1.897)	-9.605 - -0.360
Time2 x Cohort 2	2.968 (0.558)	1.759 - 4.176	7.919 (0.617)	6.552 - 9.286
Time2 x Cohort 3	2.656 (0.588)	1.365 - 3.947	5.581 (0.951)	3.227 - 7.934
Time2 x Cohort 4	6.282 (0.584)	5.001 - 7.563	6.175 (0.911)	3.935 - 8.415
Female student	1.937 (0.316)	1.298 - 2.576	-2.106 (0.304)	-2.728 - -1.484
Age	-1.685 (0.267)	-2.216 - -1.154	-1.662 (0.277)	-2.229 - -1.095
SES	0.765 (0.296)	0.087 - 1.444	0.873 (0.225)	0.391 - 1.354
Number books at home	2.343 (0.150)	2.049 - 2.638	2.027 (0.170)	1.678 - 2.375
Time x Female	0.379 (0.203)	-0.077 - 0.835	-0.662 (0.147)	-0.967 - -0.358
Time x Age	-0.353 (0.129)	-0.613 - -0.093	-0.464 (0.209)	-0.961 - 0.033
Time x SES	-0.072 (0.277)	-0.758 - 0.615	-0.119 (0.140)	-0.409 - 0.171
Time x Number books at home	-0.205 (0.106)	-0.441 - 0.031	-0.096 (0.086)	-0.277 - 0.085
Achievement Mean	0.224 (0.036)	0.152 - 0.296	0.280 (0.030)	0.220 - 0.340
Achievement SD	-0.206 (0.130)	-0.463 - 0.052	-0.141 (0.103)	-0.343 - 0.062
School SES	2.341 (0.840)	0.573 - 4.110	0.440 (0.838)	-1.302 - 2.183
Time x Achievement Mean	-0.052 (0.025)	-0.102 - -0.002	-0.043 (0.027)	-0.101 - 0.016
Time x Achievement SD	-0.134 (0.117)	-0.393 - 0.125	-0.093 (0.078)	-0.250 - 0.063
Time x School SES	-0.180 (0.461)	-1.111 - 0.750	0.672 (0.665)	-0.797 - 2.141
Female Teacher	1.550 (0.806)	-0.130 - 3.230	1.619 (0.682)	0.173 - 3.065
ITT Duration	-0.301 (0.187)	-0.672 - 0.070	-0.193 (0.252)	-0.740 - 0.354
Major	-0.856 (0.775)	-2.623 - 0.912	-1.924 (0.587)	-3.176 - -0.672
Experience	0.002 (0.027)	-0.056 - 0.060	0.031 (0.023)	-0.015 - 0.078
Time x Female Teacher	-0.763 (0.743)	-2.413 - 0.887	-0.554 (0.503)	-1.598 - 0.489
Time x ITT Duration	0.291 (0.152)	-0.018 - 0.600	0.275 (0.179)	-0.112 - 0.662
Time x Major	0.334 (0.565)	-0.936 - 1.605	1.182 (0.322)	0.548 - 1.817

	MODEL 7 LANGUAGE		MODEL 7 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Time x Experience	0.030 (0.021)	-0.014 - 0.074	-0.025 (0.021)	-0.071 - 0.020
Random Part				
Random Level 4 (School)				
Intercept	15.281 (3.559)	8.028 - 22.535	14.771 (4.626)	4.503 - 25.038
Time/Intercept	-4.001 (1.546)	-7.044 - -0.957	-4.376 (1.977)	-8.440 - -0.312
Time	4.847 (1.797)	0.904 - 8.789	5.595 (1.657)	2.176 - 9.013
Random Level 3 (Teacher)				
Intercept	19.787 (3.013)	13.164 - 26.411	21.062 (2.928)	14.752 - 27.371
Time/Intercept	-12.213 (1.966)	-16.369 - -8.057	-11.659 (2.101)	-16.156 - -7.162
Time	11.938 (1.974)	7.613 - 16.263	12.018 (2.289)	6.881 - 17.155
Random Level 2 (Student)				
Intercept	204.416 (3.308)	197.926 - 210.905	163.097 (3.244)	156.420 - 169.773
Time/Intercept	-11.249 (1.638)	-14.616 - -7.883	-1.751 (0.945)	-3.677 - 0.175
Time	5.537 (1.179)	3.116 - 7.958	0.387 (0.223)	-0.051 - 0.824
Random Level 1 (Occasion)				
Intercept	109.377 (1.322)	106.756 - 111.998	105.805 (0.890)	104.038 - 107.572
School Initial Status – Growth Correlation	-0.465		-0.481	
Teacher Initial Status – Growth Correlation	-0.795		-0.733	
Student Initial Status – Growth Correlation	-0.334		-0.220	
School ICC Initial Status	0.064		0.074	
School ICC Growth	0.217		0.311	
Teacher ICC Initial Status	0.083		0.106	
Teacher ICC Growth	0.535		0.668	
DIC	464,408		460,178	
pD	18,938		17,028	
Units: Schools	**156**		**156**	
Units: Teachers	**851**		**812**	
Units: Students	**19,704**		**19,704**	
Units: Occasions	**59,112**		**59,112**	

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions, student, teacher and school classifications. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

FIGURE 43: GROWTH CURVES BY TEACHER GENDER BASED ON MODEL 6

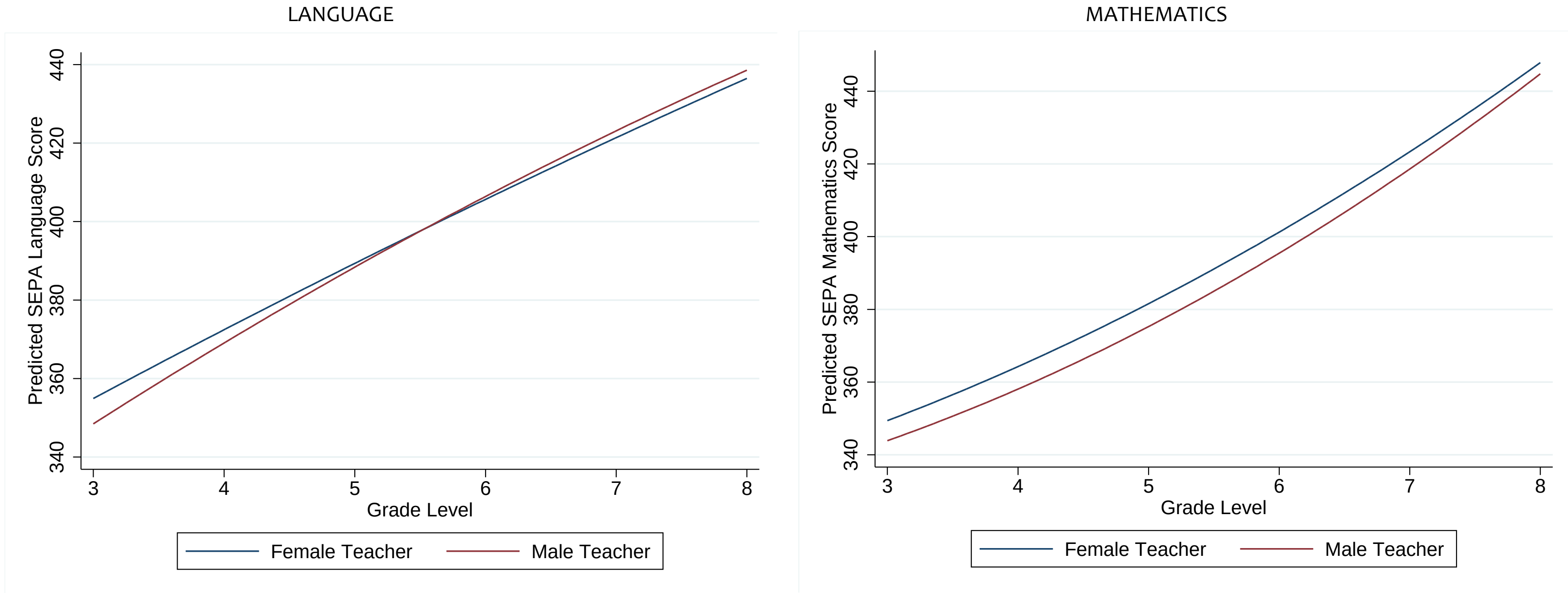


FIGURE 44: GROWTH CURVES BY INITIAL TEACHER TRAINING DURATION GROUPS BASED ON MODEL 6

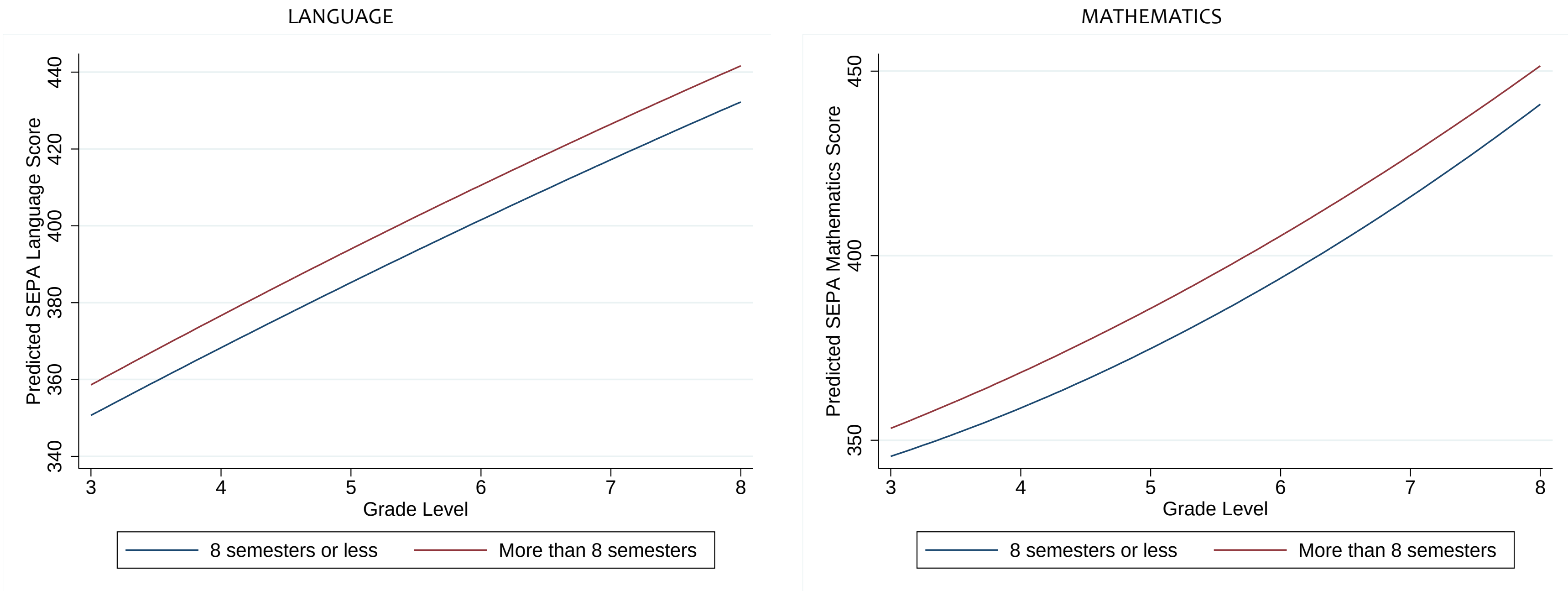


FIGURE 45: GROWTH CURVES BY TEACHER MAJOR BASED ON MODEL 6

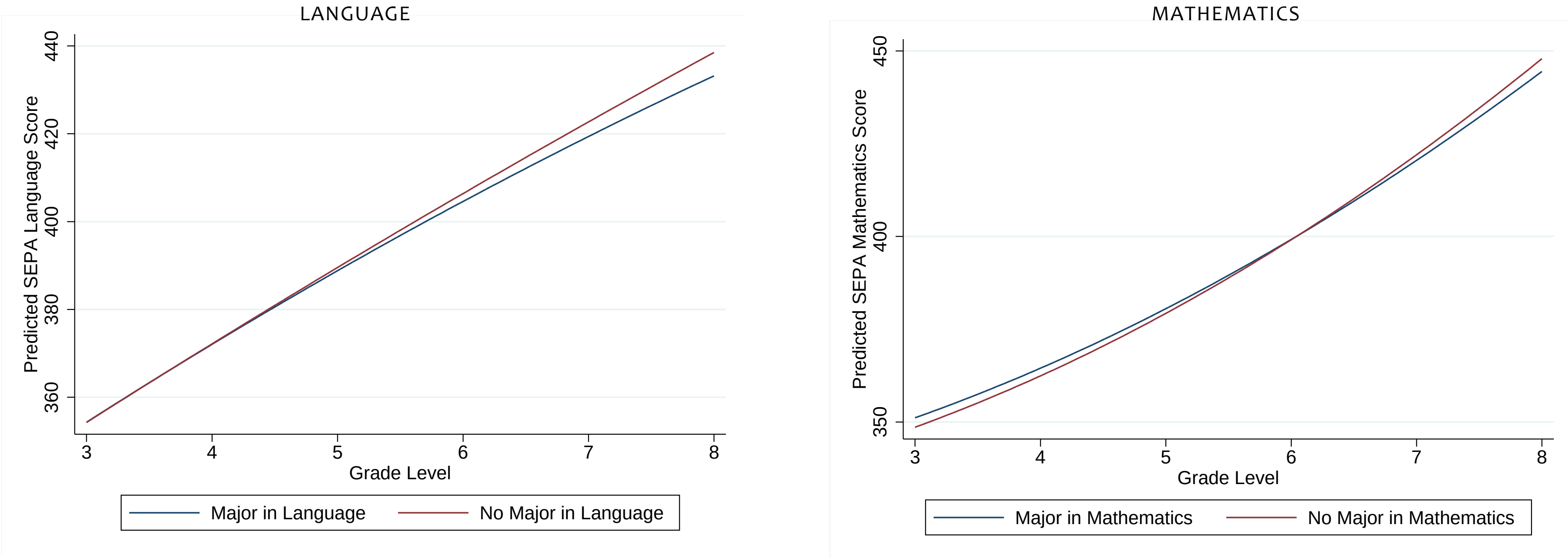
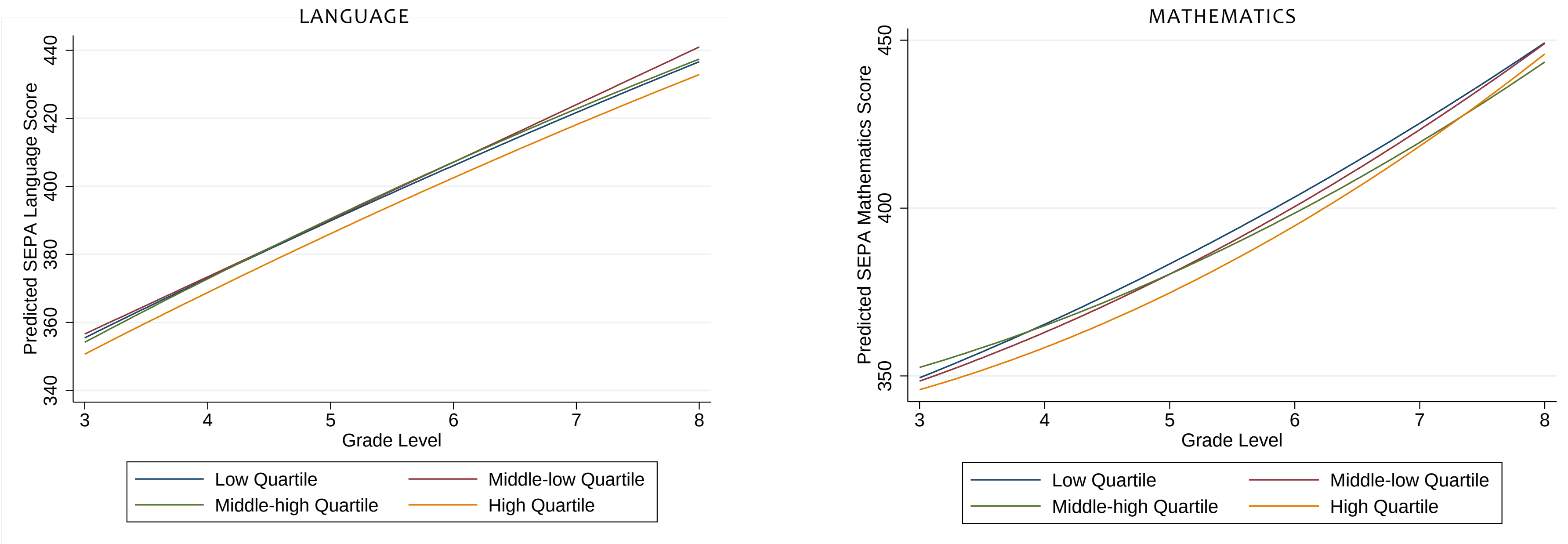


FIGURE 46: GROWTH CURVES BY TEACHERS' YEARS OF EXPERIENCE QUARTILES BASED ON MODEL 6



6.4.4 *TEACHERS' CUMULATIVE EFFECTS*

A common example of multiple membership data in educational research arises in studies of student attainment where students are often taught by more than one teacher, and so students are described as being multiple members of those teachers. In multiple membership data, lower-level units do not belong to one and only one higher-level unit. Rather, lower-level units are nested within multiple higher-level units from the same classification. The use of a cross-classified multiple membership model (Model 8) allows the depiction of teachers' cumulative effects by capturing both, the cross-classification of students and teachers and the multiple membership of students' scores to teachers.

Results from Model 8 are shown in Table 24. In both subjects, Model 8, which incorporates the multiple membership of students to teachers, fits the data better than Model 6 that does not include this multiple membership component and, therefore, assumes that teacher contributions disappear from one year to the next. This is suggested by an important reduction on DIC values from Model 6 to Model 8 (ΔDIC = -388 in language, and ΔDIC = -147 in mathematics). These results indicate that previous teachers continue to influence student achievement scores later on in time.

Given the limitation of the three time points available in this study, it is only possible to look at growth across two academic years for each student and results reveal teacher effects remain cumulative in this time scale in both language and mathematics subjects for the multiple cohorts covered from Grade 3 to Grade 8.

TABLE 24: FULL MODEL 8 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE AND MATHEMATICS

	MODEL 8 LANGUAGE		MODEL 8 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Fixed Part				
Intercept	348.202 (0.766)	346.662 - 349.742	345.087 (0.818)	343.448 - 346.726
Time	33.989 (0.953)	31.848 - 36.129	30.762 (1.257)	27.720 - 33.803
Time2	-7.411 (0.463)	-8.476 - -6.347	-7.805 (0.737)	-9.678 - -5.932
Cohort 2	24.967 (0.598)	23.784 - 26.150	22.880 (0.677)	21.484 - 24.276
Cohort 3	37.247 (0.697)	35.850 - 38.645	34.681 (0.656)	33.376 - 35.987
Cohort 4	55.285 (0.680)	53.930 - 56.639	57.304 (0.639)	56.042 - 58.566
Time x Cohort 2	-9.273 (0.980)	-11.332 - -7.215	-15.479 (1.243)	-18.297 - -12.662
Time x Cohort 3	-9.034 (0.893)	-10.854 - -7.214	-4.362 (1.337)	-7.474 - -1.250
Time x Cohort 4	-13.240 (0.930)	-15.140 - -11.340	-3.649 (1.522)	-7.292 - -0.006
Time2 x Cohort 2	2.639 (0.440)	1.727 - 3.551	8.123 (0.675)	6.528 - 9.718
Time2 x Cohort 3	2.237 (0.408)	1.426 - 3.048	4.962 (0.727)	3.235 - 6.689
Time2 x Cohort 4	5.971 (0.486)	4.964 - 6.977	5.679 (0.799)	3.747 - 7.611
Female student	1.836 (0.323)	1.181 - 2.492	-2.242 (0.295)	-2.840 - -1.645
Age	-1.660 (0.268)	-2.193 - -1.128	-1.689 (0.275)	-2.252 - -1.126
SES	0.742 (0.290)	0.080 - 1.404	0.915 (0.226)	0.431 - 1.400
Number books at home	2.315 (0.153)	2.014 - 2.616	1.998 (0.168)	1.653 - 2.342
Time x Female	0.375 (0.207)	-0.090 - 0.841	-0.625 (0.147)	-0.928 - -0.322
Time x Age	-0.342 (0.132)	-0.608 - -0.076	-0.454 (0.211)	-0.956 - 0.048
Time x SES	-0.032 (0.265)	-0.683 - 0.620	-0.143 (0.135)	-0.419 - 0.134
Time x Number books at home	-0.200 (0.104)	-0.429 - 0.029	-0.096 (0.086)	-0.276 - 0.085
Achievement Mean	0.188 (0.048)	0.093 - 0.284	0.231 (0.041)	0.149 - 0.314
Achievement SD	-0.209 (0.192)	-0.597 - 0.179	-0.017 (0.161)	-0.338 - 0.305
School SES	2.216 (1.050)	0.049 - 4.382	1.352 (1.028)	-0.693 - 3.397
Time x Achievement Mean	-0.044 (0.026)	-0.095 - 0.006	-0.016 (0.029)	-0.079 - 0.047
Time x Achievement SD	-0.165 (0.138)	-0.465 - 0.134	-0.111 (0.088)	-0.285 - 0.063
Time x School SES	-0.137 (0.539)	-1.219 - 0.946	0.164 (0.697)	-1.304 - 1.632
Random Part				

	MODEL 8 LANGUAGE		MODEL 8 MATHEMATICS	
	M (SE)	95% CI	M (SE)	95% CI
Random Level 4 (School)				
Intercept	33.872 (6.100)	21.755 - 45.990	38.614 (8.845)	19.602 - 57.625
Time/Intercept	-14.472 (3.557)	-21.886 - -7.059	-16.793 (3.782)	-24.607 - -8.98
Time	10.606 (3.271)	3.145 - 18.067	11.591 (2.193)	7.229 - 15.952
Random Level 3 (Teacher)				
Intercept	33.020 (5.039)	21.571 - 44.468	45.086 (8.507)	24.369 - 65.803
Time/Intercept	-14.956 (2.755)	-21.471 - -8.440	-20.347 (4.664)	-32.081 - -8.614
Time	8.190 (1.547)	4.487 - 11.893	10.452 (2.528)	4.032 - 16.872
Random Level 2 (Student)				
Intercept	201.946 (3.279)	195.511 - 208.382	160.808 (3.521)	153.377 - 168.24
Time/Intercept	-11.807 (1.640)	-15.232 - -8.382	-2.623 (0.993)	-4.649 - -0.598
Time	5.598 (1.116)	3.250 - 7.946	0.780 (0.271)	0.247 - 1.312
Random Level 1 (Occasion)				
Intercept	108.641 (1.202)	106.270 - 111.012	104.724 (1.162)	102.240 - 107.207
School Initial Status – Growth Correlation	-0.764		-0.794	
Teacher Initial Status – Growth Correlation	-0.909		-0.937	
Student Initial Status – Growth Correlation	-0.351		-0.234	
School ICC Initial Status	0.126		0.158	
School ICC Growth	0.435		0.508	
Teacher ICC Initial Status	0.123		0.184	
Teacher ICC Growth	0.336		0.458	
DIC	**464,008**		**460,010**	
pD	**18,935**		**17,115**	
Units: Schools	**156**		**156**	
Units: Teachers	**851**		**812**	
Units: Students	**19,704**		**19,704**	
Units: Occasions	**59,112**		**59,112**	

6.5 Conclusions

This chapter examined teacher effects on student achievement trajectories with more advanced methods than in most previous research. First, three data points have been used. Previous studies, in contrast, rely largely on cross-sectional data or two-data-points designs, which either confound age and cohort effects or contain a very limited source of intra-individual variability to study change. Research has shown that with three data points the straight-line achievement growth model can be evaluated and the precision of the parameter estimates improves (Raudenbush and Liu, 2000; Rogosa, Brand and Zimowski, 1982). The three-time point longitudinal design provides a considerable source of intra-individual variability, and the statistical models soundly distinguish within-student, between-student, between-teacher and between-school variability in a partially crossed design.

Secondly, the effect of teachers on student achievement trajectories was estimated by means of crossed-random multiple membership effects, longitudinal data with an accelerated design, and multilevel models. These techniques properly distinguish intra- and inter-individual variability. Furthermore, cross-random-effects models soundly account for within-school variability in partially crossed designs. From the literature reviewed it has been found that most methodological examinations have neglected the crossed nature of the data when modelling teacher effects.

Thirdly, this study advances the field methodologically by demonstrating the combined use of state-of-the-art techniques, namely, accelerated longitudinal designs, growth curve approaches, and cross-classified multiple membership models. Thus, the research is methodologically innovative and advances prior research. While this is the main original contribution to the literature, theoretical implications are also important.

Consistent with previous research, the present analyses show that both teacher and school effects are substantially larger than previously reported due to the use of more appropriate models and measurement. In addition, the research confirms that teacher effects at primary level exceed school effects. It is clear that educational effects are larger when achievement progress over time, rather than achievement status, is studied, which confirms that teachers but also to a lesser extent schools make an important contribution to student achievement growth. Also, teacher effects are not highly consistent across subjects and the contribution of teachers to student achievement growth seems to accumulate over time, at least in the short time scale studied (two academic years), and this holds for multiple cohorts from Grade 3 to Grade 8.

What is new and intriguing about these results is that they reveal that teacher effects seem to be larger in an emerging economy such as Chile than in post-industrialised countries, which can be indicative of more unequal access to high-quality teaching in the former. However, the specific mechanisms generating teacher effects were not investigated in this study, which was concerned with measuring the magnitude and accumulation of teacher effects, and the applied methodological advance of measurement.

CHAPTER 7: DISCUSSION AND CONCLUSIONS

The present chapter is structured as follows: First, the findings of the different phases of the research project are brought together and discussed in light of the literature, and the contributions of the study to the field of educational effectiveness research are highlighted. Then, the main strengths and limitations of the study are discussed and suggestions for future research are made. The chapter ends with some implications for Chilean education policy.

7.1 Main Results and Contributions to the Field of Educational Effectiveness Research

The present study was designed to investigate student achievement trajectories and educational effects in Chile. To this end, secondary data obtained from different administrative reporting systems and assessment projects, containing information on a wide range of students, teachers and schools, were merged and analysed using advanced multilevel modelling. In this section, the substantive findings of the thesis are linked to educational effectiveness research, with a focus on the contribution that the conclusions make and how they add to our cumulative knowledge base.

7.1.1 Student Achievement Trajectories

As shown in Chapter 4, a quadratic growth curve gave better fit to the data than a linear trend, indicating non-linear upward growth on student achievement from grades 3 to 8 of primary school, for both language and mathematics. In addition, individual students were found to differ substantially in both their achievement onset and their rate of development over time. Student characteristics that were found to explain variations in student achievement levels were students' gender, age, socio-economic status and number of books at home.

Consistently with the literature, the analyses show that students of higher SES families perform better in language and mathematics than their lower SES peers. Achievement differences related to family SES are explained by parental education and family wealth influences. What is new and intriguing from the results is that they reveal that the achievement gap among SES groups tends to remain stable during the primary school years. These findings are fairly robust with respect to different model specifications, and there is no evidence of ceiling limits imposed by the tests on the achievement levels or growth of higher SES students. Even within the private non-subsidised school sector, the percentage of students achieving the maximum score is negligible. Nonetheless, this trend is not in accordance with what most studies anticipate (e.g., Alexander, Entwisle and Olson, 2001; 2007; Caro, McDonald and Willms, 2009; Downey, von Hippel and Broh, 2004), namely, that the gap between students of high- and low-SES families widens with age (i.e., the Matthew effect). However, other studies have also found that achievement differences among students of varying socio-economic backgrounds remain invariant, or even narrow, during primary school (Caro and Lehmann, 2009; Wilkins and Ma, 2002).

The specific mechanisms generating this pattern of stable inequalities were not investigated in this study, which is more concerned with measuring the longitudinal trend in the SES achievement gap. That said, the stability of the SES gap in primary school years can be understood in light of the distinctive characteristics of the Chilean educational system, where social stratification and segmentation between schools might differentiate student trajectories at an earlier stage. It is possible that, as previous studies have indicated (Alexander et al. 2001; Caro et al. 2009a), the SES gap emerges early in life, before the ages captured in this study, and remains stable for most of the primary school years.

In the literature, student gender has also been considered an important predictor of achievement levels. This study adds to the existent research on gender differences in achievement growth rates, which is currently inconclusive and lacking evidence on the primary school level. It was found that girls' language achievement levels were higher than boys', and this difference remained stable over the primary school ages. The growth curves for mathematics achievement were also a function of student gender. While most of the previous studies find no gender differences in the rate of growth in mathematics or, if they do, they are found in the secondary school level, in the present study boys achieve slightly higher than girls, and their trajectory shows a stronger acceleration throughout primary school. This is an interesting result that requires further exploration, particularly by looking at gender differences among specific groups of students (e.g., across different performance levels) and by investigating the mechanisms that might cause gender differences in achievement trajectories.

7.1.2 School Effects

Chapter 5 addressed four themes in the measurement of school effects: magnitude, consistency, compositional effects and differential effectiveness. School effects in two cognitive outcomes, language and mathematics achievement, were modelled through a three-level accelerated growth curve model. It was found that there are considerable school differences regarding achievement status and growth in both subjects. Furthermore, the use of growth trajectories showed once again that school differences with respect to the growth of their students are bigger than usually demonstrated in studies with cross-sectional and two time points longitudinal designs. This study, thus, underscores the importance of longitudinal studies in educational effectiveness research as cross-sectional studies may fail to discover important educational effects. Furthermore, the sizes of the school effects found here are somewhat larger than those found in previous studies using alike model specifications for similar outcomes in the other national contexts (e.g. Guldemond and Bosker, 2009; Lenkeit, 2012; Palardy, 2010; Rowan, Correnti and Miller, 2002). This is in line with previous literature, suggesting that larger school effects are found in emerging economies when compared to post-industrialised countries.

School-level variables that predict achievement status are school overall achievement and SES. The school compositional effects detected have to do with differences in achievement levels already manifest at intake (i.e., Grade 3 of primary school). As in previous studies (Belfi et al., 2013; Guldemond and Bosker, 2009; Luyten, Schildkamp and Folmer, 2009), school compositional effects on students' growth were not found. This study calculated overall achievement and SES compositional effects using external school variables (not produced through the aggregation of the student variables in the models), which helps to reduce the bias in the estimation of compositional effects introduced by measurement error (Televantou et al., 2015). The results also show that sector differences are due to differences in student intake and not in school effectiveness.

As in previous research (Luyten, 1994; Marks, 2015; Sammons, Mortimore and Thomas, 1996; Scheerens and Bosker, 1997; Thomas et al., 1997b; Willms and Raudenbush, 1989), moderate consistency (r = 0.489) was found between school effects on language and school effects on mathematics achievement levels. More importantly, some degree of consistency was also found regarding school effects on growth in the two outcomes (r = 0.425). In addition, a negative association was found between the intercept and the slope of the two cognitive outcomes, indicating that the growth in achievement was larger in schools attended by students with lower achievement levels at the end of Grade 3. Importantly, it was checked that this pattern is not an artifact of ceiling limits on achievement levels or growth of advantaged students.

Finally, school effects in Chile were shown to vary across student groups. More specifically, the gender gap in language achievement growth differed somewhat from school to school, as found in previous research (Strand, 2010). Most studies on differential effectiveness have found none or only modest differential effects with regard to student socioeconomic status (Kyriakides, 2004, Sammons et al., 1993, Strand, 2010, Thomas, 2001, Thomas et al., 1997a). This study shows small differential effects across students of different socioeconomic status in both subjects, indicating that some schools are better than others at reducing the overall patterns of difference related to students' backgrounds.

7.1.3 Teacher Effects

As with schools, several properties of teacher effects were investigated in Chapter 6, namely, their magnitude, consistency, predictors and cumulativeness. One clear implication of these analyses is that researchers need to move beyond the use of both covariate adjustment models and annual gains models if they want to estimate the overall magnitudes of teacher effects on growth in student achievement. A promising strategy shown in this study is to use a cross-classified random effects model, as Raudenbush (1995) discusses. The analyses reported here suggest that cross-classified random effects models lead to findings of larger *d*-type teacher effects. The magnitudes found are even larger than reported in the few previous studies that have also accounted for crossed grouping factors in the data when estimating teacher effects (Palardy, 2010;

Raudenbush and Bryk, 2002; Rowan, Correnti and Miller, 2002). Furthermore, the study contributes with evidence to a key debate in educational effectiveness research, as it shows that teacher effects outweigh school effects (Luyten, 2003).

Findings on the consistency of teacher effects across academic subjects were also presented. Using a cross-classified accelerated growth model, it was found that the same teacher was not highly consistently effective across different academic subjects, confirming the evidence from previous studies in other contexts (Campbell et al. 2003; Muijs et al. 2005; Rowan, Correnti and Miller, 2002; Sammons et al. 2007). This indicates that teachers are likely to develop higher competence in some subjects.

The teacher-to-teacher differences in effects on student achievement growth imply that some students make less academic progress than they would otherwise be expected to make, simply by being placed in ineffective classrooms. This suggests that the important problem is not simply to demonstrate that differences in effectiveness exist among teachers, but rather to explain why these differences occur and to improve teaching effectiveness broadly. Although most of the input teacher variables tested in this study (i.e., teacher gender, duration of initial teacher training and years of teaching experience) did not account for an important proportion of the variation on teacher effects, holding a major in mathematics was found to be predictive of student achievement growth in primary school. This variable can be seen as an indirect measure of teachers' knowledge of the content being taught. This result is in line with prior research, which has found positive effects of measures of teachers' knowledge on student achievement (e.g., Greenwald, Hedges and Laine, 1996). While not addressed in this study, the international evidence also suggests that classroom process variables, if well measured, hold promise for explaining differences in teacher effectiveness.

Finally, a multiple membership component was incorporated into the cross-classified random effects model by using a weighting scheme where the membership weights were proportional to the time each student spent with a given teacher. This allowed for depicting teachers' cumulative effects over time. The cross-classified multiple membership model fitted the data significantly better than the cross-classified model that does not consider the multiple membership of students to teachers, and therefore assumes that teacher contributions disappear at the end of each year, suggesting that the effects of teachers accumulate over time as previous teachers continue to influence student achievement scores in subsequent years.

7.1.4 Contributions to the field of Educational Effectiveness

This study addresses three important gaps in the literature as it, firstly, explores educational effects in the context of an emerging economy using appropriate measures and specifications, secondly, contributes further evidence on the properties of school and teacher effects on student achievement growth (i.e., magnitude, consistency, predictors and differential effects) and, thirdly, advances the field methodologically by demonstrating the combined use of accelerated longitudinal designs, growth curve approaches, and cross-classified and multiple membership models. These contributions to the literature are explained below.

7.1.4.1 Exploring educational effects in Chile using adequate EER models

The study tested and found support for several aspects of existing EER models, such as the Dynamic Model, in the context of a developing country: Chile. Furthermore, the conceptual model proposed in Chapter 2 expands previous EER models by emphasising on, and depicting in greater detail, the effects of time, which in this study refer to change in conditions for different cohorts and change in students' achievement as they progress through school.

As mentioned in previous chapters, the existing studies on educational effectiveness in Chile have all used cross-sectional data or longitudinal data with, at most, two time points. The design of this study is longitudinal, collecting data from the same children over three time points. This is the first educational effectiveness study with more than two time points of longitudinal data in the country, and the second in the Latin-American region. Furthermore, the study is the first in the region that annually matches students with their teachers and models the relationships between students and their successive classroom settings.

Modelling school and teacher effects on student achievement growth represents a clear improvement when compared to previous measures available in the Chilean and Latin

American contexts. Thus, this thesis expands the field as it presents the first study in the Chilean context that makes use of annually collected longitudinal data on student achievement linked to extensive student, teacher and school information. Making use of these rich longitudinal data, the study appropriately disentangles the contribution of school and teachers on student outcomes from part of the variance introduced by measurement error and, with this, provides better estimates of educational effects than those obtained by previous local research.

7.1.4.2 RESEARCHING THE PROPERTIES OF SCHOOL AND TEACHER EFFECTS ON STUDENT ACHIEVEMENT GROWTH

As shown in the conceptual framework presented in Chapter 2, the study incorporates and provides evidence on important components and dimensions of educational effectiveness (i.e., children's cognitive growth over time, teacher and school effects, compositional effects, consistency across subjects and differential effectiveness across student groups).

The analysis of variation in students' educational outcomes, and the way various sources of influence help to shape students' learning and developmental outcomes over time, is well supported by multilevel modelling, which has become a powerful tool for EER (Luyten and Sammons, 2010; Rumberger and Palardy, 2004). Furthermore, the accelerated longitudinal cohort design used in the current study lends itself naturally to value-added modeling, because the data consists of repeated measures on students who cross teachers within schools over time. The hierarchical, crossed-level, value-added-effects model applied can be conceptualised as the joining of two separate hierarchical models, the first of which is a two-level model for individual growth in achievement over time, and the second of which is a two-level model of the value added that each teacher and school contributes to student learning in each particular year.

Thus, this research is unique as it simultaneously investigates several properties of school and teacher effects using a longitudinal approach that has proved to be a more sensible strategy for detecting educational effects. This comprehensive examination provides a better insight into the complex dynamics of educational effectiveness.

7.1.4.3 *USING AND COMBINING ADVANCED LONGITUDINAL DESIGNS AND MULTILEVEL MODELS*

The study investigates educational effects in Chile with more advanced methods than in the past. First, the study uses three data points. Previous studies, in contrast, rely largely on cross-sectional data or longitudinal data with two time-points that either confound grade and cohort effects or contain a very limited source of intra-individual variability to study change in achievement. Research has shown that with three data points the linear and curvilinear growth models can be evaluated and the precision of the parameter estimates improved (Raudenbush and Liu, 2000; Rogosa, Brand and Zimowski, 1982). Furthermore, the linking of four overlapping cohorts of data in an accelerated longitudinal model adds robustness to the study of student achievement trajectories (Miyazaki and Raudenbush, 2000).

Secondly, student achievement trajectories and educational effects from Grades 3 to 8 are estimated by means of advanced multilevel models (i.e., growth curve, cross-classified and multiple membership models). These techniques properly distinguish intra- and inter-individual variability. Furthermore, cross-classified models soundly account for within-school variability in partially crossed designs (Beretvas, 2008). The longitudinal data in this study are partially crossed because students can be taught by multiple teachers over time, and therefore the multilevel structure breaks down. Cross-classified models efficiently handle large and partially crossed datasets and thus nicely correspond to the data. According to literature reviewed, previous examinations have usually neglected the crossed nature of the data when modelling teacher effects.

The accelerated three time point longitudinal design provides a considerable source of intra-individual variability, and the statistical models soundly distinguish within-student, between-student, between-teacher and between-school variability in a partially crossed design. Thus, this methodological examination is innovative and advances prior research.

7.2 STRENGTHS AND LIMITATIONS OF THE STUDY

There are noteworthy strengths in the study undertaken. As mentioned above, the sound design and innovative methodological approach of this research project advances the field of educational effectiveness in a number of ways. The main strengths of the present study are:

7.2.1 *A COMPREHENSIVE SAMPLE*

The research design of this study is quite robust and makes use of extensive data collected from a number of sources, which allowed data validation. The study incorporates a wide range of schools, teachers and students, and benefits from unprecedentedly rich longitudinal data on the Chilean system. In addition, the sample sizes are large which is beneficial in terms of statistical power to detect the effects investigated.

As explained in Chapter 3, the final sample has not been randomly selected but includes a large number of schools that have voluntarily decided to participate in the SEPA Assessment Project. Therefore, it is not possible to confirm that the same findings would hold in a nationally representative data or in other educational systems. The somewhat biased composition of the sample in this study, over-representing students from more advantageous home environments and attending high-SES, private and large schools, demands caution in the interpretation and generalisation of findings. Furthermore, the study only considers students that progressed without delay through the three years of primary education in which they were assessed. Underrepresentation of retained students in the analytic sample may introduce a bias in the results; this represents a limitation of the study. Nonetheless, there are no better data sources in the Chilean context to inform the issues addressed in this study. Other data sets available in Chile contain fewer measurement points and lack information on school, teachers or students.

7.2.2 *MULTIPLE MEASURES OF EDUCATIONAL ACHIEVEMENT*

With regards to the measures of educational achievement used, the study benefits from data on the two most frequently explored educational outcomes: language and mathematics attainment. This facilitated important comparisons and analyses of consistency across academic subjects that add robustness to the study of student achievement trajectories and educational effects.

The analyses conducted were not based on data collected specifically for the purposes of this study. External teams of experts designed the instruments and collected these secondary data. This potential limitation was addressed by carrying out analyses of reliability and validity of the measures in order to assess their quality, which was found to be more than satisfactory.

Unfortunately, several student, teacher and school variables, whose importance have been emphasised in contemporary conceptual models of educational effectiveness (Creemers, 1994; Creemers and Kyriakides, 2008), were not available in the secondary data used in this study. For example, teacher process variables, that can be indicative of teaching effectiveness, were not available. Finally, since the dependent variables refer only to language and mathematics knowledge, the study has not been able to monitor pupils' progress in other aspects of the school curriculum, nor did it examine educational effectiveness in relation to other important goals of education, such as the development of metacognitive skills or social and affective outcomes.

7.2.3 A Longitudinal and Multiple-Cohort Design

Collecting longitudinal data in educational research enabled the investigation of school and teacher effects on the development of student outcomes. In an era of increased accountability of schools and teachers, they should not be held responsible for the type of students they attract (i.e., school intake differences) but only for the changes that students make that can be attributed to school factors. Longitudinal data allow us to model the changes in student outcomes over time, and, thus, these types of data are needed in order to measure school and teacher effects adequately.

It is important to note that, as the SEPA project is a system based on 12-month learning, the estimation of school and teacher effects based on these data implicitly incorporates learning that occurs over the summer. It has been shown in the literature that children's summer learning rates are very unequal (much more unequal than they are during the school year) and the children who make greater summer gains tend to be relatively more affluent and have higher initial scores (Cooper et al., 1996; Downey, von Hippel and Broh, 2004; Entwisle and Alexander, 1992). Thus, although annual learning rates are far more valid than achievement levels, 12-month learning is still not a perfect measure of educational effects (Palardy, 2010; Palardy and Peng, 2015). However, modelling school and teacher effects on student achievement growth represents a clear improvement when compared to previous measures available in the Chilean and Latin-American contexts, and the effect of differential summer learning rates was expected to be weakened by the use of several control variables in the growth models implemented, such as student and school SES and a proxy for cultural capital (number of books at home).

Together with using a longitudinal design, the criteria for satisfactory inference in studies of educational effectiveness have often included the analysis of multiple cohorts (Goldstein, 1997; Gray, 1995; Sammons, Thomas and Mortimore, 1997). However, longitudinal multiple-cohort data is still infrequent in educational effectiveness research. By using an accelerated longitudinal design, and statistical models that appropriately fit this data structure, the robustness of the study is increased, as results are confirmed across both grade levels and cohorts. Furthermore, this accelerated longitudinal design permits the study of a wider range of school years (six out of the eight years of primary school in Chile) than a traditional three-year longitudinal study.

In this regard, a potential limitation of the study is the limited number of overlapping time points between cohorts, as adjacent cohorts share two overlapping time points. It has been noted that, when designing accelerated longitudinal studies, care needs to be taken to plan overlaps between cohorts which are adequate to test for the hypothesized developmental patterns: the more complex the predicted pattern of growth, the greater the number of overlapping time points is needed (Anderson, 1993). The study would have, therefore, benefited from data on these cohorts over more time points.

7.2.4 A Multilevel Approach

In this study, the challenges imposed by the complex hierarchical structure of the longitudinal data were dealt with by applying advanced multilevel models. The multilevel models implemented correspond to contextualised value-added models, which have become popular and widely used in both EER and accountability systems. In Chapters 4 to 6 these models accommodated both multiple measures of the same student (via growth curve modelling), and multiple cohorts of students (using accelerated longitudinal modelling). In Chapter 6, a cross-classified accelerated growth model for the analysis of teacher effects was fitted as some students change teachers across years, and therefore, student outcomes do not follow the traditional nested design of hierarchical models. In addition, the hypothesis of teacher cumulative effects was tested by comparing the fit of the latter model to that of a cross-classified multiple membership accelerated growth model. The complexity of these models matches the system and phenomena being studied to a larger extent than models used in previous research.

7.2.5 Appropriate Missing Data Treatment

As the study combines information from several secondary sources, it is affected by the quality of these data. A limitation of the present study is data loss due to missing values and attrition. While, at the school level, data from the SIMCE Assessment System, the SIGE Recording System and the MINEDUC School Enrolment Recording System are generally complete, difficulties appear at the teacher and student levels. The inexact identification of students based on the National Identification Number in the SEPA and the SIMCE Assessment Systems led to a considerable loss of information when matching the SEPA and the SIMCE data sets to extract student background information from the latter. Moreover, missing data in longitudinal research is particularly problematic due to attrition (Hill and Goldstein, 1998). In the case of the present study, attrition is produced by students changing schools and by schools leaving and joining the SEPA project.

These issues were appropriately dealt with by conducting extensive analyses of the patterns and mechanisms of missing data, and by implementing a suitable strategy for handling missing data, namely, multilevel multiple imputation. This modern missing data method makes better use of the observed information, increases robustness to non-ignorable missingness and improves estimation precision (Schafer and Graham, 2002).

Furthermore, the complex hierarchical structure of the data posed challenges to multiple imputation that were solved by using an appropriate strategy: applying an unrestricted two-level imputation model with test scores in wide format, students as level 1, and schools as level 2, and linking together the language and mathematics databases so the imputation of data in the language data set would benefit from information on mathematics test scores as auxiliary variables, and vice versa.

In conclusion, as with all research, the present study is not free of limitations. The main restrictions, discussed above, were imposed by the current availability and quality of secondary data in Chile and can be summarised as five different issues: the representativeness of the sample, the unspecified differential summer learning rates, the outcome measures available, and the limited overlapping time points between cohorts and missing data. The following section presents suggestions for future research arising from this study.

7.3 Recommendations for Future Research

In the process of conducting this research, gaps in the knowledge base and potential extensions to the study undertaken have been identified. Recommendations for future studies are summarised in this section.

7.3.1 Focus on Early Learning Experiences

The present study detected that substantive variation in student attainment and achievement gaps related to gender and SES are already well established by the time students are in Grade 3 of primary school. Therefore, it would be of great interest to analyse students' cognitive outcomes in earlier stages and from a longitudinal approach in order to detect when these gaps start to appear, and to what extent early learning experiences contribute to explaining education inequalities once children enter school and later on in life.

The importance of early childhood education on students' later achievement has been well documented in various studies across many countries (e.g., Barnett, 2008; Camilli et al., 2010; Sylva et al., 2010). The evidence has shown that children from affluent families are more likely to experience a more positive home learning environment and preschool experience compared to children from disadvantaged backgrounds, which, in turn, leads to differences in the cognitive development of individuals.

Chile has significantly expanded early education coverage for children from low-income background (UNESCO, 2015b). However, there is still room for improvement in terms of participation levels, that remain below the OECD average (OECD, 2015a). Furthermore, previous research has found significant differences in the quality of early childhood education, with highly heterogeneous preschool provision (Cortázar, 2015) and teacher performance (Treviño, Toledo and Gempp, 2013), which is likely to contribute to later differences in educational outcomes.

Research that contributes with regional evidence to the direct and mediating role of pre-school in SES and gender gaps, and that informs the policy debate on the importance of early interventions for reducing education inequalities, is greatly needed.

7.3.2 *Follow-up Study on the Secondary School Level*

As data from new waves of the SEPA Assessment Project become available, the present study has the potential to be extended to include data from students in the sample as they go on to secondary school. Methodologically, this would provide even richer longitudinal data with increased number of overlapping time points between cohorts, allowing the modelling of more complex random effects models of predicted patterns of achievement growth. Substantially, this would provide evidence on student achievement trajectories, attainment gaps and educational effects in a critical educational phase prior to the key transition from school to tertiary education and work.

7.3.3 *Estimating the Rate of Decay of Teacher Effects*

A more specific issue that could be addressed in future research is the estimation of the rate of decay of teacher effects. The cross-classified multiple membership accelerated growth model fitted in Chapter 6 assumes that teacher effects persist undampened into the future. The validity of this assumption has not been fully explored in the literature, and while there is evidence that teacher effects are long-lasting, there is considerable reason to conjecture that a teacher's effect will dampen over time as students grow and are exposed to other teachers and other learning experiences (McCaffrey et al., 2004). For example, we might think that the order in which teachers have taught is important. In particular we might want to allow the final teacher assigned to have a greater influence than earlier teachers. Alternatively, we might reason that each year being taught by a teacher during primary schooling is more important than the last. This issue can be explored by comparing alternative teacher weighting schemes, as part of a sensitivity analysis for the multiple membership model. Furthermore, this approach would also allow the study of the variation in the long-term effects of teachers across subjects and student groups.

7.3.4 Exploring the Effect of Teacher and School Process Variables

The fact that a sizeable percentage of between-school and between-teacher variance remains after controlling for the available student, school and teacher variables indicates that malleable educational conditions rather than merely student selection factors are likely to account for differences in student achievement growth. Research including variables such as school climate, classroom environment and teacher instructional processes could help to disentangle these large educational effects.

This study, as with most research on predictors of educational effects in the region, has been driven by input (e.g., teacher characteristics) and context (e.g., school composition) variables. There are few studies that examine the relationship between value-added scores and teaching processes. However, what happens inside the classroom (e.g., teacher practices and interactions between the teacher and their students) has been regarded as the black box in educational research and has proved to be more predictive of educational outcomes than input and context variables (Currie and Neidell, 2007).

Indeed, educational effectiveness research has drawn attention to the centrality of classroom processes in determining schools' overall academic effectiveness. It has been argued that the quality of teaching and expectations has the most significant role to play in fostering students' learning and progress (Creemers, 1994; Sammons, Hillman and Mortimore, 1995; Scheerens and Bosker, 1997). Given this, school processes, including leadership and climate, remain influential because they provide the overall framework within which teachers and classrooms operate.

In Chile, there is need of more evidence on the relationship between teacher effects and observational measures of teacher quality in different contexts of schooling, integrating the input, process and context paradigms in educational effectiveness research. These measures can be obtained by conducting classroom observations using systematic observation protocols, such as the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro and Hamre, 2008) and the International System for Teacher Observation and Feedback (ISTOF) (Teddlie et al., 2006). The integration of observational data in the study of teacher effectiveness will provide implications for teaching, particularly in relation to what constitutes good practice in terms of promoting better outcomes for students in Chile. This approach would also allow researchers to investigate whether Chilean teachers provide equally, less or more instructional support than practitioners in other countries.

7.3.5 *Systematising Available Latin-American Longitudinal Data*

This thesis was carried out to investigate educational inequalities and effectiveness in Chile. No one study on its own can set a complete picture of all the factors affecting student achievement trajectories. However, the results of this study add weight to the claim that schools, and particularly teachers, make an important contribution to children's educational attainment.

The generally large between-school variation in developing countries makes for a particularly interesting context for educational effectiveness research. However, in Latin America, most research still relies on cross-sectional data. Only in recent years have other countries in the region conducted longitudinal educational studies (i.e., Cuba, Brazil, Mexico and Peru), but these longitudinal data have, so far, been under-utilised. Examples of such databases are the GERES study, a longitudinal research programme focused on elementary school effectiveness in Brazil (Brooke and Bonamino, 2011), and the Young Lives Project, a long-term international study investigating the changing nature of childhood poverty in Peru and other countries (Bourdillon and Boyden, 2014). These rich sources of data have not been used to compare educational effects in the region. Analysing and comparing the evidence on student achievement trajectories and educational effects arising from these longitudinal databases would certainly contribute to the advancement of the field and to education policy in the region.

Having discussed the suggestions for future research arising from this study, the following section discusses the implications for education policy in Chile.

7.4 Implications for Chilean Education Policy

This study aims to impact policymakers by contributing to the debate on education effectiveness and inequality, as it produced new evidence on these issues with state-of-the art methods. The main issues that are informed by the evidence presented are discussed below.

7.4.1 Adding Complexity to School and Teacher Evaluation Systems

The evaluation of school and teacher effects based on analyses of student attainment data is increasingly being used as a key strand of school inspection and teacher evaluation systems across the world. In Chile, school and teacher performance measures are being discussed as part of a new high-stakes accountability system. This study is very timely, given that in recent years Chile has given way to a major policy shift towards the introduction of new laws and accountability mechanisms (Santiago et al., 2013; Treviño and Donoso, 2010).

The approval by the Congress of the Ley General de Educación (LGE) put in place in 2013 the 'Sistema Nacional de Aseguramiento de la Calidad de la Educación' (National System of Assurance of the Quality in Education) that encompasses the creation of two new bodies: the Superintendence of Education, and the Quality Agency of Education. The latter, a local version of the Office for Standards in Education (OFSTED), is mandated to not just classify schools in four performance groups (good, satisfactory, fair and poor) based on school SIMCE assessment results and other data, but also will be implementing high-stakes consequences on the basis of this classification (ranging from organisational interventions up to school closure) for those schools that do not meet the required standards.

The present research has highlighted the importance of schools and teachers on student achievement growth, a finding consistent with previous international studies. However, this study has also confirmed that in Chile there is a stronger than average relationship between student academic performance, socio-economic background and school composition. Thus, it could be seen as highly problematic that the new institutional assessment framework does not intend to use more sophisticated and sensitive approaches to evaluating school performance that could statistically adjust for these factors, such as contextualised value-added measures. The implication is that the annual categorisation of schools will not take proper account of contextual or compositional effects at the level of the school, suggesting that the collective effects of group processes on individuals' performances are not important. Particularly, the decision to judge school performance based on their raw test results would be a way of rewarding school selection and segregation. The claim that something more sophisticated than raw 'league tables' is needed in order to compare schools on a more equitable basis has been strongly associated with contextualised value-added approaches (Schagen and Hutchison, 2003) as they offer a more rigorous approach for levelling the playing field among schools.

While it is not reasonable to expect that a new accountability system would match the complexity of the models implemented in this study, the results presented add to the ongoing debate about fairer ways to judge the effectiveness of schools and teachers. This new evidence can inform and potentially improve the future credibility and legitimacy of school and teacher evaluation systems under implementation in Chile and in other contexts.

7.4.2 *The Misleading Private Advantage*

The present study concluded that school sector was not a significant predictor of student achievement growth, after controlling for student background and school composition. In other words, there were no statistical differences in terms of effectiveness between private subsidised schools and municipal/public schools. This is an important finding for policy makers, practitioners and parents in the Chilean context given generalised claims, based on the publication of SIMCE raw scores, that private subsidised schools are better than public schools.

Public schools in Chile do show lower achievement levels than private schools, but their students are making as good progress as their peers in the private sector, and they are not given credit for this under the generalised use of an unadjusted raw score measures. Against the vision that public schools are inefficient, the results showed that public schools are typically helping their students make progress, and that the introduction of a contextualised value-added measure in the accountability system is particularly relevant for schools working in disadvantaged contexts.

Furthermore, this indicates that parental pressure through school choice in Chile has not had the effect on school effectiveness claimed by the voucher supporters, and that privatisation and competition, the key drivers of the policy reforms in the last decades, have not led to a general improvement of outcomes in the system.

7.4.3 *Intervening School Composition*

The results presented contribute with evidence of school compositional effects across subject areas and characterise the levels of education inequality (within schools and between schools). This is informative for policymakers, as school composition in Chile is partially driven by system level properties, such as school selectivity, school choice and urban segregation (Elacqua, 2012; McEwan, Urquiola, & Vegas, 2008; Valenzuela, Bellei, & Ríos, 2014).

Indeed, the effect of student and school SES on students' achievement levels should be taken seriously by policymakers. The present study has underlined the strong links between SES and students' academic achievement and the disadvantages produced by attending schools with intake from a low SES area. In order to ensure equal chances for all students to succeed at school, this crucial issue that hinders the equity of the Chilean education system should be addressed. Interventions focusing on decreasing the effect of socio-economic status, such as monitoring the school-average SES, could be reviewed by policymakers and adjusted to the Chilean education system.

7.4.4 *Focus on Pre-school Education*

The study shows that, by the time students reach the end of Grade 3, differences in both their achievement levels and rate of development over time are already established. Furthermore, SES gaps are in place and remain constant over the course of the primary school years.

Improving both the coverage and the quality of pre-school education, particularly for children from socially disadvantaged backgrounds, could help to reduce the SES gap in further education levels, as the long-term positive impact of good quality nursery attendance on achievement is well documented.

7.4.5 *Improving Gender Equality in Education*

This study provided evidence on how gender differences develop over time in language and mathematics achievement. Male students were found to perform better and progress faster in mathematics than female students. These findings should be taken into consideration when exploring ways to increase students' achievement and reduce gender inequalities in education.

Gender parity in access to quality education can be a key factor in fostering development and reducing inequality. However, regional and socio-economic differences can promote or hinder education of boys or girls by imposing unequal expectations regarding their capacities and future roles. A way of moving towards gender equity in the school system is to encourage parents and teachers to raise self-confidence and motivation among girls to pursue interests in mathematics. Also, curricula, teaching material and training policies should be reviewed to avoid gender stereotyping.

In addition, the magnitude and direction of the gender gap in language achievement growth were found to vary across schools, as there were indications of school differential effectiveness in this respect. It would be of interest to policy makers then to look in depth at the features and practices of those more equitable schools.

7.4.6 Subject Specialisation in Initial Teacher Training

Initial teacher training has been an area neglected by educational policy during the last decades, and it has only recently been brought to the public debate. The creation of a Panel of Experts in 2010, commissioned by the Ministry of Education to analyse the state of the teacher training system and recommend reforms, are signs that the issue is gaining momentum. The report emanated from this commission highlighted the lack of research on the features of teacher education that have an impact on teachers' professional performance, and acknowledges the issue that the information about teachers that could enable this area of inquiry is currently scarce (Panel of Experts for Quality Education, 2010).

The results of this study clearly indicate that teacher effects in the primary school level are sizeable. When looking at the predictors of these large effects, neither the duration of their initial training nor their years of teaching experience help to explain the variance on their students' achievement growth. However, holding a major in mathematics was predictive of teacher effectiveness in the subject. Furthermore, teachers were not found to be equally effective across subjects. While more research is needed in this area, this is a first indication that actions towards improving teachers' levels of subject specialisation can promote student achievement progress, and that differences in terms of subject-specific training of teacher education programmes should be monitored.

In summary, as discussed in this chapter, the results presented in this thesis provide a basis for future studies and for fruitful discussions and comparisons on issues related to academic achievement in Chile and internationally. This is expected to assist in the promotion of excellence and equity in the Chilean education system.

REFERENCES

Ai, X. (2002) Gender differences in growth in mathematics achievement: Three-level longitudinal and multilevel analyses of individual, home, and school influences. *Mathematical Thinking and Learning,* 4 (1), 1-22.

Aitkin, M. and Longford, N. (1986) Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General),* 149 (1), 1-43.

Alexander, K. L., Entwisle, D. R. and Olson, L. S. (2001) Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis,* 23 (2), 171-191.

Alexander, K. L., Entwisle, D. R. and Olson, L. S. (2007) Lasting consequences of the summer learning gap. *American Sociological Review,* 72 (2), 167-180.

Alvarado, M., Cabezas, G., Falck, D. and Ortega, M. E. (2012) La evaluación docente y sus instrumentos: Discriminación del desempeño docente y asociación con los resultados de los estudiantes [The teacher evaluation system and its instruments: Teacher performance prediction and association with students' results]. Santiago: MINEDUC-PNUD.

Amrein-Beardsley, A. (2008) Methodological concerns about the education value-added assessment system. *Educational Researcher,* 37 (2), 65-75.

Anand, P., Mizala, A. and Repetto, A. (2009) Using school scholarships to estimate the effect of private education on the academic achievement of low-income students in Chile. *Economics of Education Review,* 28 (3), 370-381.

Anderson, E. R. (1993) Analyzing change in short-term longitudinal research using cohort-sequential designs. *Journal of Consulting and Clinical Psychology,* 61 (6), 929-940.

Antoniou, P. (2012) The short- and long-term effects of secondary schools upon students academic success and development. *Educational Research and Evaluation,* 18 (7), 621-639.

Anumendem, N., De Fraine, B., Onghena, P. and Van Damme, J. (2013) The impact of coding time on the estimation of school effects. *Quality & Quantity,* 47 (2), 1021-1040.

Asparouhov, T. and Muthén, B. (2010) *Multiple imputation with MPLUS.* Los Angeles, CA: Muthén & Muthén.

Avalos, B. and Matus, C. (2010) *La formación inicial docente en Chile. Informe nacional del estudio internacional IEA TEDS-M [Initial teacher training in Chile. National report of the international study IEA TEDS-M].* Santiago: Gobierno de Chile.

Ballou, D., Sanders, W. and Wright, P. (2004) Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics,* 29 (1), 37-65.

Barnett, S. W. (2008) *Preschool education and its lasting effects: Research and policy implications.* New Jersey: National Institute for Early Education Research Rutgers, The State University of New Jersey.

Belfi, B., Goos, M., Pinxten, M., Verhaeghe, J. P., Gielen, S., De Fraine, B. and Van Damme, J. (2013) Inequality in language achievement growth? An investigation into the impact of pupil socio-ethnic background and school socio-ethnic composition. *British Educational Research Journal,* 40 (5), 820-846.

Bell, R. Q. (1953) Convergence: An accelerated longitudinal approach. *Child Development,* 24 (2)**,** 145-152.

Bellei, C. (2005) The private-public school controversy: The case of Chile. Cambridge, MA: John F. Kennedy School of Government, Harvard University.

Bellei, C., Muñoz, G., Pérez, L. M. and Raczynski, D. (2003) Escuelas efectivas en sectores de pobreza [Effective schools in poor areas]. *In:* R. Hevia (ed.) *La educación en Chile, hoy [Education in Chile today].* Santiago: Ediciones Universidad Diego Portales.

Beretvas, N. (2008) Cross-classified random effects models. *In:* A. A. O'Connell and D. B. McCoach (eds.) *Multilevel modeling of educational data.* Charlotte, NC: Information Age Publishing.

Beretvas, S. N. (2011) Cross-classified and multiple membership models. *In:* J. J. Hox and J. K. Roberts (eds.) *Handbook of advanced multilevel analysis.* New York: Routledge.

Biancarosa, G., Bryk, A. and Dexter, E. (2010) Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal,* 111 (1)**,** 7-34.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A. and Curran, P. J. (2004) The role of coding time in estimating and interpreting growth curve models. *Psychological methods,* 9 (1)**,** 30-52.

Bingham, R. D., Heywood, J. S. and White, S. B. (1991) Evaluating schools and teachers based on student performance. *Evaluation Review,* 15 (2)**,** 191-218.

Bock, R. D., Wolfe, R. and Fisher, T. (1996) Review and analysis of the Tennessee value-added assessment system. Nashville: Office of Education Accountability.

Bourdillon, M. and Boyden, J. (2014) *Growing up in poverty: Findings from young lives.* Oxford: Palgrave Macmillan.

Bravo, D., Contreras, D. and Sanhueza, C. 1999. Rendimiento educacional, desigualdad, y brecha de desempeño privado/público: Chile 1982-1997 [Educational attainment, inequality and the private/public attainment gap: Chile 1982-1997]. Santiago: Universidad de Chile.

Breen, R. and Goldthorpe, J. H. (1997) Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society,* 9 (3)**,** 274-305.

Bressoux, P. and Bianco, M. (2004) Long-term teacher effects on pupils' learning gains. *Oxford Review of Education,* 30 (3)**,** 327-345.

Brooke, N. and Bonamino, A. (2011) *Geres 2005: Razões e resultados de uma pesquisa longitudinal sobre a eficácia escolar [Geres 2005: Reasons and results of a longitudinal study on school effectiveness].* Rio de Janeiro: WalPrint.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J. and Wisenbaker, J. (1979) *School social systems and student achievement: Schools can make a difference.* New York: Praeger.

Brophy, J. and Good, T. (1986) Teacher behavior and student achievement. *In:* M. Wittrock (ed.) *Handbook of research on teaching.* New York: Macmillan.

Browne, W. J. (2012) *MCMC estimation in MLWin, v2.26.* Bristol: Centre for Multilevel Modelling, University of Bristol.

Browne, W. J. and Draper, D. (2006) A comparison of bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis,* 1 (3)**,** 473-549.

Browne, W. J., Goldstein, H. and Rasbash, J. (2001) Multiple membership multiple classification (mmmc) models. *Statistical Modelling,* 1 (2)**,** 103-124.

Brunner, J. J. and Uribe, D. (2007) *Mercados universitarios: El nuevo escenario de la educación superior [University markets: The new higher education scenario].* Santiago: Universidad Diego Portales.

Bryk, A. S. and Raudenbush, S. W. (1987) Application of hierarchical linear models to assessing change. *Psychological Bulletin,* 101 (1), 147-158.

Buchmann, C. (2002) Measuring family background in international studies of education: Conceptual issues and methodological challenges. In: A. C. Porter and A. Gamoran (eds.) *Methodological advances in cross-national surveys of educational achievement* Washington, DC: National Academy Press.

Caldas, S. J. and Bankston, C. L. (1999) Multilevel examination of student, school, and district- level effects on academic achievement. *The Journal of Educational Research,* 93 (2), 91–100.

Camilli, G., Vargas, S., Ryan, S. and Barnett, W. S. (2010) Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record,* 112 (3), 579-620.

Campbell, R. J., Kyriakides, L., Muijs, R. D. and Robinson, W. (2003) Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education,* 29 (3), 347-362.

Campbell, R. J., Kyriakides, L., Muijs, R. D. and Robinson, W. (2004) *Assessing teacher effectiveness: Developing a differentiated model.* London: RoutledgeFalmer.

Cariola, L., Cares, G. and Lagos, E. (2009) *¿Qué nos dice PISA sobre la educación de los jóvenes en chile? Nuevos análisis y perspectivas sobre los resultados en PISA 2006 [What does PISA tell us about young people's education in Chile? New analyses and perspectives on PISA 2006 results].* Santiago: Unidad de Currículum y Evaluación, Ministerio de Educación.

Caro, D. and Lenkeit, J. (2012) An analytical approach to study educational inequalities: 10 hypothesis tests in PIRLS 2006. *International Journal of Research & Method in Education,* 35 (1), 3-30.

Caro, D., McDonald, T. and Willms, J. D. (2009) A longitudinal analysis of math achievement gaps associated with socio-economic status: How do they change from childhood to adolescence? *Canadian Journal of Education,* 32 (3), 558-590.

Caro, D. H. and Cortés, D. (2012) Measuring family socioeconomic status: An illustration using data from PIRLS 2006. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments,* 5, 9-33.

Caro, D. H. and Lehmann, R. (2009) Achievement inequalities in hamburg schools: How do they change as students get older? *School Effectiveness and School Improvement,* 20 (4), 407-431.

Caro, D. H., Lenkeit, J., Lehmann, R. and Schwippert, K. (2009) The role of academic achievement growth in school track recommendations. *Studies In Educational Evaluation,* 35 (4), 183-192.

Cervini, R. A. (2009) Class, school, municipal, and state effects on mathematics achievement in Argentina: A multilevel analysis. *School Effectiveness and School Improvement,* 20 (3), 319-340.

Cheung, C. and Rudowicz, E. (2003) Academic outcomes of ability grouping among junior high school students in Hong Kong. *The Journal of Educational Research,* 96 (4), 241–254.

Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L. (2006) Teacher-student matching and the assessment of teacher effectiveness. Cambridge: National Bureau of Economic Research.
Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L. (2007a) How and why do teacher credentials matter for student achievement? Cambridge, MA: National Center for Analysis of Longitudinal Data in Education Research.
Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L. (2007b) Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review,* 26 (6)**,** 673-682.
Cole, N. S. (1997) *The ETS gender study: How females and males perform in educational settings.* Princeton: Educational Testing Service.
Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A., Weinfield, F. D. and York, R. L. (1966) *Equality of educational opportunity.* Washington: U.S. Government Printing Office.
Collins, L. M., Schafer, J. L. and Kam, C.-M. (2001) A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods,* 6 (4)**,** 330-351.
Condron, J. (2007) Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems,* 54 (1)**,** 139-160.
Consejo Nacional de Educación (2015) Proceso matrícula 2015: Evolución de carreras de educación [Enrolment process 2015: Evolution of education programmes]. Santiago: Consejo Nacional de Educación.
Cook, T. D. and Campbell, D. T. (1979) *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin.
Cooper, H., Nye, B., Charlton, K., Lindsay, J. and Greathouse, S. (1996) The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research,* 66 (3)**,** 227-268.
Cortázar, A. (2015) Long-term effects of public early childhood education on academic achievement in Chile. *Early Childhood Research Quarterly,* 32**,** 13-22
Cox, C. (2010) Educational inequality in Latin America. Patterns, policies and issues. *In:* P. Attewell and K. S. Newman (eds.) *Growing gaps. Educational inequality around the world.* Oxford: Oxford University Press.
Cox, C., Meckes, L. and Bascopé, M. (2010) La institucionalidad formadora de profesores en Chile en la década del 2000: Velocidad del mercado y parsimonia de las políticas [Initial teacher education institutions in Chile between 2000 and 2010: Market speed and policies' parsimony]. *Revista Pensamiento Educativo,* 46-47**,** 205-245.
Creemers, B. P. M. (1994) *The effective classroom.* London: Cassell.
Creemers, B. P. M. and Kyriakides, L. (2006) Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement,* 17 (3)**,** 347-366.
Creemers, B. P. M. and Kyriakides, L. (2008) *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools.* London: Routledge.
Creemers, B. P. M., Kyriakides, L. and Sammons, P. (2010) *Methodological advances in educational effectiveness research.* London: Routledge.

Creemers, B. P. M. and Reezigt, G. J. (1992) *Evaluation of educational effectiveness.* Groningen: ICO.

Currie, J. and Neidell, M. (2007) Getting inside the "black box" of head start quality: What matters and what doesn't. *Economics of Education Review,* 26 (1), 83–99.

Darling-Hammond, L. (2000) Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives,* 8 (1), 1-50.

De Fraine, B., Damme, J. V., Landeghem, G. V., Opdenakker, M. C. and Onghena, P. (2003) The effect of schools and classes on language achievement. *British Educational Research Journal,* 29 (6), 841-859.

De Fraine, B., Damme, J. V. and Onghena, P. (2007) A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology,* 32 (1), 132-150.

De Fraine, B., Landeghem, G., Damme, J. and Onghena, P. (2005) An analysis of wellbeing in secondary school with multilevel growth curve models and multilevel multivariate models. *Quality & Quantity: International Journal of Methodology,* 39 (3), 297-316.

De Fraine, B., Van Damme, J. and Onghena, P. (2002) Accountability of schools and teachers: What should be taken into account? *European Educational Research Journal,* 1 (3), 403-428.

De Haan, A., Elbers, E., Hoofs, H. and Leseman, P. (2012) Targeted versus mixed preschools and kindergartens: Effects of class composition and teacher-managed activities on disadvantaged children's emergent academic skills. *School Effectiveness and School Improvement,* 24 (2), 1-18.

Donoso, S. (2008) Políticas de perfeccionamiento de los docentes en Chile 1990-2005: Silencios y proyecciones [Teacher improvement policies in Chile 1990-2005: Silences and proyections]. In: C. Bellei, D. Contreras and J. P. Valenzuela (eds.) *La agenda pendiente en educación [The pending agenda in education].* Santiago: Universidad de Chile-UNICEF.

Downey, D., von Hippel, P. and Broh, B. (2004) Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review,* 69 (5), 613-635.

Draper, D. (2008) Bayesian multilevel analysis and MCMC. *In:* J. d. Leeuw and E. Meijer (eds.) *Handbook of multilevel analysis.* New York: Springer.

Dumay, X., Coe, R. and Anumendem, D. N. (2013) Stability over time of different methods of estimating school performance. *School Effectiveness and School Improvement,* 25 (1), 1-19.

Duncan, S. C., Duncan, T. and Hops, H. (1996) Analysis of longitudinal data within accelerated longitudinal designs. *Psychological methods,* 1 (3), 236–248.

Duncan, S. C. and Duncan, T. E. (2012) Accelerated longitudinal designs. In: B. Laursen, T. D. Little and N. A. Card (eds.) *Handbook of developmental research methods.* New York: Guilford Publications.

Elacqua, G. (2012) The impact of school choice and public policy on segregation: Evidence from Chile. *International Journal of Educational Development,* 32 (3), 444-453.

Elacqua, G., Schneider, M. and Buckley, J. (2006) School choice in Chile: Is it class or the classroom? *Journal of Policy Analysis and Management,* 25 (3), 577-601.

Enders, C. K. (2006) A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine Research,* 68 (3), 427-436.

Enders, C. K. and Tofighi, D. (2007) Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological methods,* 12 (2), 121-138.

Entwisle, D. and Alexander, K. (1992) Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review,* 57 (1), 72-84.

Erikson, R. and Rudolphi, F. (2010) Change in social selection to upper secondary school, primary and secondary effects in Sweden. *European Sociological Review,* 26 (3), 291-305.

Farrell, J. P. and Oliveira, J. B. A. (1993) *Teachers in developing countries: Improving effectiveness and managing costs.* Washington: World Bank.

Fielding, A. and Goldstein, H. (2006) Cross-classified and multiple membership structures in multilevel models: An introduction and review. London: Department for Education and Skills.

Flores, C. (2008) Segregacao residencial e resultados educacionais na cidade de Santiago do Chile [Residential segregation and educational outcomes in the city of Santiago, Chile]. In: R. Katzman and L. Queiroz (eds.) *A cidade contra a escola? [The city against the school].* Rio de Janeiro: Letra Capital.

Foley, B. and Goldstein, H. (2012) *Measuring success. League tables in the public sector.* London: The British Academy.

Fuller, B. and Clarke, P. (1994) Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research,* 64 (1), 119-157.

Gamoran, A. (1992) The variable effects of high school tracking. *American Sociological Review,* 57 (6), 812-828.

García-Huidobro, J. (2009) Institucionalidad educativa chilena: Análisis desde una perspectiva democrática [Educational institutionality in Chile: Analysis from a democratic perspective]. Santiago: Universidad Alberto Hurtado.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian data analysis.* New York: Chapman & Hall.

Goldhaber, D. D. and Brewer, D. J. (2000) Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis,* 22 (2), 129-145.

Goldstein, H. (1987) Multilevel covariance component models. *Biometrika,* 74 (2), 430-431.

Goldstein, H. (1997) Methods in school effectiveness research. *School Effectiveness and School Improvement,* 8 (4), 369-395.

Goldstein, H. (2011) *Multilevel statistical models.* Hoboken: Wiley.

Goldstein, H. and Healy, M. J. R. (1995) The graphical presentation of a collection of means. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 158 (1), 175-177.

Goldstein, H., Simon, B. and Brendon, M. (2007) Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 170 (4), 941-954.

Goldthorpe, J. H. (1996) Class analysis and the reorientation of class theory: The case of persisting differentials in educational attainment. *The British Journal of Sociology,* 47 (3), 481-505.

Gray, J. (1995) *A multi-level analysis of school improvement: Changes in schools' performance over time.* Lisse: Swets & Zeitlinger.

Greenwald, R., Hedges, L. V. and Laine, R. D. (1996) The effect of school resources on student achievement. *Review of Educational Research,* 66 (3), 361-396.

Guldemond, H. and Bosker, R. J. (2009) School effects on students' progress, a dynamic perspective. *School Effectiveness and School Improvement,* 20 (2), 255-268.

Guo, G. (1998) The timing of the influences of cumulative poverty on children's cognitive ability and achievement. *Social Forces,* 77 (1), 257-288.

Haller, E. J. and Davis, S. A. (1981) Teacher perceptions, parental social status, and grouping for reading instruction. *Sociology of Education,* 54 (3), 162-174.

Hamaker, E. L., van Hattum, P., Kuiper, R. M. and Hoijtink, H. J. A. (2011) Model selection based on information criteria in multilevel modeling. In: J. Hox and J. K. Roberts (eds.) *Handbook of advanced multilevel analysis.* New York: Routledge.

Hanushek, E. A. (1989) The impact of differential expenditures on school performance. *Educational Researcher,* 18 (4), 45-62.

Hanushek, E. A. (1994) Money might matter somewhere: A response to hedges, laine, and greenwald. *Educational Researcher,* 23 (4), 5-8.

Hanushek, E. A. (1995) Interpreting recent research on schooling in developing countries. *The World Bank Research Observer,* 10 (2), 227-246.

Hanushek, E. A. (1997) Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis,* 19 (2), 141-164.

Hanushek, E. A., Kain, J., O'Brien, D. and Rivkin, S. G. (2005) The market for teacher quality. Cambridge: The National Bureau of Economic Research.

Hanushek, E. A., Kain, J. F., Markman, J. M. and Rivkin, S. G. (2003) Does peer ability affect student achievement? *Journal of Applied Econometrics,* 18 (5), 527-544.

Harker, R. and Tymms, P. (2004) The effects of student composition on school outcomes. *School Effectiveness and School Improvement,* 15 (2), 177-199.

Hattie, J. (2009) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Hauser, R. M. (1994) Measuring socioeconomic status in studies of child development. *Child Development,* 65 (6), 1541-1545.

Heck, R. H. (2009) Teacher effectiveness and student achievement: Investigating a multilevel cross-classified model. *Journal of Educational Administration,* 47 (2), 227-249.

Herzog, T., Scheuren, F. and Winkler, W. (2007) *Data quality and record linkage techniques.* New York: Springer.

Heyneman, S. P. and Loxley, W. A. (1982) Influences on academic achievement across high and low income countries: A re-analysis of IEA data. *Sociology of Education,* 55 (1), 13-21.

Hill, P. W. and Goldstein, H. (1998) Multilevel modeling of educational data with cross-classification and missing identification for units. *Journal of Educational and Behavioral Statistics,* 23 (2), 117-128.

Hill, P. W. and Rowe, K. J. (1996) Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement,* 7 (1), 1-34.

Hill, P. W. and Rowe, K. J. (1998) Modelling student progress in studies of educational effectiveness. *School Effectiveness and School Improvement,* 9 (3)**,** 310-333.

Hill, P. W. and Russell, V. J. (1999) Systemic, whole-school reform of the middle years of schooling. *In:* R. J. Bosker, B. P. M. Creemers and S. Stringfield (eds.) *Enhancing educational excellence, equity and efficiency.* Amsterdam: Springer.

Hox, J. J. (2010) *Multilevel analysis: Techniques and applications.* London: Routledge.

Hsieh, C. T. and Urquiola, M. (2006) The effects of generalized school choice on achievement and stratification: Evidence from Chile's school voucher program. *Journal of Public Economics,* 90 (8)**,** 1477–1503.

IBM 2011. IBM SPSS missing values 20. IBM Corporation.

Ingvarson, L. C., Schwille, J., Tatto, M. T., Rowley, G., Peck, R. and Senk, S. L. (2013) *An analysis of teacher education context, structure, and quality-assurance arrangements in TEDS-M countries: Findings from the iea teacher education and development study in mathematics (TEDS-M).* Amsterdam: Association for the Evaluation of Educational Achievement (IEA).

Inter-American Development Bank (2014) Urgency and possibility. Results of PRIDI: A first initiative to create regionally comparative data on child development in four Latin American countries. New York: Inter-American Development Bank.

Jackman, S. (2009) *Bayesian analysis for the social sciences.* Chichester: Wiley.

Jencks, C., Smith, M., Acland, H., Bane, M. S., Cohen, D., Gintis, H., Heyns, B. and Michelson, S. (1973) *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Jesson, D. and Gray, J. (1991) Slants on slopes: Using multi-level models to investigate differential school effectiveness and its impact on pupils' examination results. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice,* 2 (3)**,** 230-247.

Jordan, H. R., Mendro, R. L. and Weerasinghe, D. 1997. *Teacher effect on longitudinal student achievement.* Paper presented at the CREATE Annual Meeting, Indianapolis (IN), USA, July. Unpublished.

Kang, C., Park, C. and Lee, M. J. (2007) Effects of ability mixing in high school on adulthood earnings: Quasi-experimental evidence from South Korea. *Journal of Population Economics,* 20 (2)**,** 269-297.

Kerckhoff, A. C. (1993) *Diverging pathways: Social structure and career deflections.* New York: Cambridge University Press.

Kiplinger, V. L. 2004. Longitudinal study of student growth in reading, writing, mathematics achievement in thirty-six Colorado public school district. Phase II. Final report. Available from: http://matrix10.d20.co.edu/lssg/phase2.html. [Accessed 17 August 2015]

Koedel, C. and Betts, J. R. 2008. Test-score ceiling effects and value-added measures of school quality. *JSM Proceedings, Social Statistics Section.* Alexandria, VA: American Statistical Association.

Kolaweski-Jones, L. and Duncan, G. J. (1999) The structure of achievement and behavior across middle childhood. *Child Development,* 70 (4)**,** 930-943.

Kreft, I. G. G., De Leew, J. and Aiken, L. S. (1995) The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research,* 30**,** 1-21.

Kukla-Acevedo, S., Streams, M. and Toma, E. F. (2009) Evaluation of teacher preparation programs: A reality show in Kentucky. Lexington: University of Kentucky.

Kupermintz, H. (2003) Teacher effects and teacher effectiveness: A validity investigation of the tennessee value added assessment system. *Educational Evaluation and Policy Analysis,* 25 (3), 287-298.

Kyriakides, L. (2004) Differential school effectiveness in relation to sex and social class: Some implications for policy evaluation. *Educational Research and Evaluation,* 10 (2), 141-161.

Kyriakides, L. and Creemers, B. P. M. (2008) A longitudinal study on the stability over time of school and teacher effects on student outcomes. *Oxford Review of Education,* 34 (5), 521-545.

Lara, B., Mizala, A. and Repetto, A. (2010) Una mirada a la efectividad de los profesores en Chile [An overview of teacher effectiveness in Chile]. *Estudios Públicos,* 120, 147-182.

Lara, B., Mizala, A. and Repetto, A. (2011) The effectiveness of private voucher education: Evidence from structural school switches. *Educational Evaluation and Policy Analysis,* 33 (2), 119-137.

Leahey, E. and Guo, G. (2001) Gender differences in mathematical trajectories. *Social Forces,* 80 (2), 713-732.

Leckie, G. (2009) The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 172 (3), 537-554.

Leckie, G. and Charlton, C. (2012) Runmlwin: A program to run the MLWin multilevel modelling software from within Stata. *Journal of statistical software,* 52 (11), 1-40.

Lenkeit, J. (2012) Effectiveness measures for cross-sectional studies: A comparison of value-added models and contextualised attainment models. *School Effectiveness and School Improvement,* 24 (1), 39-63.

León, M., Manzi, J. and Paredes, R. (2009) Calidad docente y rendimiento escolar en Chile: Evaluando la evaluación [Teacher quality and student achievement in Chile: Evaluating the evaluation]. Santiago: Pontificia Universidad Católica de Chile.

Levine, D. and Lezotte, L. (1990) Unusually effective schools: A review and analysis of research and practice. *School Effectiveness and School Improvement,* 1 (3), 221-224.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical analysis with missing data.* Hoboken: Wiley.

Lockwood, J. R. and McCaffrey, D. F. (2009) Exploring student-teacher interactions in longitudinal achievement data. *Education Finance and Policy,* 4 (4), 439-467.

Lockwood, J. R., McCaffrey, D. F., Mariano, L. T. and Setodji, C. (2007) Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics,* 32 (2), 125-150.

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012) *The bugs book: A practical introduction to bayesian analysis.* Boca Raton: Chapman and Hall-CRC.

Luo, W. and Kwok, O. (2012) The consequences of ignoring individuals' mobility in multilevel growth models. *Journal of Educational and Behavioral Statistics,* 37 (1), 31-56.

Luyten, H. (1994) Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research,* 21 (2), 197-216.

Luyten, H. (1998) School effectiveness and student achievement, consistent across subjects? Evidence from Dutch elementary and secondary education. *Educational Research and Evaluation,* 4 (4)**,** 281-306.

Luyten, H. (2003) The size of school effects compared to teacher effects: An overview of the research literature. *School Effectiveness and School Improvement,* 14 (1)**,** 31-51.

Luyten, H. and Sammons, P. (2010) Multilevel modelling. *In:* B. P. M. Creemers, L. Kyriakides and P. Sammons (eds.) *Methodological advances in educational effectiveness research.* New York: Routledge.

Luyten, H., Schildkamp, K. and Folmer, E. (2009) Cognitive development in dutch primary education, the impact of individual background and classroom composition. *Educational Research and Evaluation,* 15 (3)**,** 265-283.

Ma, X. (1999) Gender differences in growth in mathematical skills during secondary grades: A growth model analysis. *Alberta Journal of Educational Research,* 45 (4)**,** 448-466.

Ma, X. (2008) Within-school gender gaps in reading mathematics, and science literacy. *Comparative Education Review,* 52 (3)**,** 437-460.

Manzi, J., Lacerna, P., Meckes, L., Ramos, I. and Ortega, L. (2012) ¿Qué características de la formación inicial de los docentes se asocian a mayores avances en su aprendizaje de conocimientos disciplinarios? [What are the characteristics of teacher initial training associated with teachers' progress in disciplinary knowledge?]. Santiago: Ministerio de Educación.

Manzi, J., San Martín, E. and Van Bellegem, S. (2014) School system evaluation by value added analysis under endogeneity. *Psychometrika,* 79 (1)**,** 130-153.

Marks, G. N. (2008) Gender differences in the effects of socioeconomic background: Recent cross-national evidence. *International Sociology,* 23 (6)**,** 845-863.

Marks, G. N. (2015) The size, stability, and consistency of school effects: Evidence from Victoria. *School Effectiveness and School Improvement,* 26 (3)**,** 397-414.

Martineau, J. A. (2006) Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics,* 31 (1)**,** 35-62.

Martínez, J. F. (2012) Consequences of omitting the classroom in multilevel models of schooling: An illustration using opportunity to learn and reading achievement. *School Effectiveness and School Improvement,* 23 (3)**,** 305-326.

Marzano, R. J., Pickering, D. and Pollock, J. E. (2001) Classroom instruction that works: Research-based strategies for increasing student achievement. Alexandria, Va: Association for Supervision and Curriculum Development.

Matear, A. (2007) Equity in education in Chile: The tensions between policy and practice. *International Journal of Educational Development,* 27 (1)**,** 101-113.

May, H., Supovitz, J. A. and Perda, D. (2004) A longitudinal study of the impact of America's choice on student performance in Rochester, New York, 1998–2003. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A. and Hamilton, L. (2004) Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics,* 29 (1)**,** 67-101.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M. and Hamilton, L. (2003) Evaluating value-added models for teacher accountability. Santa Monica: RAND Corporation.

McCoach, D. B., O'Connell, A. A., Reis, S. M. and Levitt, H. A. (2006) Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology,* 98 (1), 14-28.

McEwan, P. J. (2001) The effectiveness of public, catholic, and non-religious private schools in Chile's voucher system. *Education Economics,* 9 (2), 103-128.

McEwan, P. J. and Carnoy, M. (2000) The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis,* 22 (3), 213-239.

Mella, O. (2003) 12 años de reforma educacional en Chile. Algunas consideraciones en torno a sus efectos para reducir la inequidad [12 years of educational reform in Chile. Some considerations about its effects on reducing inequity]. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación,* 1 (1), 1-12.

Meyers, J. L. and Beretvas, S. N. (2006) The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research,* 41 (4), 473-497.

Microdatos - University of Chile (2006) Encuesta longitudinal de docentes. Primera versión: Informe final [Longitudinal survey of teachers. First version: Final report]. Santiago: Microdatos, University of Chile.

MIDE UC (2015) *El proyecto SEPA [The SEPA project].* [Online]. Available from: http://www.sepauc.cl/ [Accessed 4 October 2015].

MINEDUC (2012) *Sistema de medición de la calidad de la educación (SIMCE) [The educational quality measuring system (SIMCE)].* [Online]. Available from: http://www.simce.cl/ [Accessed 18 June 2012].

MINEDUC (2015a) Estadísticas de la educación en Chile 2014 [Educational statistics in Chile 2014]. Santiago: Ministerio de Educación.

MINEDUC (2015b) Resultados de la evaluación Inicia 2014 [Inicia assessment results 2014]. Santiago: Ministerio de Educación.

Miyazaki, Y. and Raudenbush, S. W. (2000) Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological methods,* 5 (1), 44-63.

Mizala, A. and Romaguera, P. (1998) Desempeño escolar y elección de colegios: La experiencia chilena [Student achievement and school choice: The Chilean experience]. Santiago: Universidad de Chile.

Mizala, A. and Romaguera, P. (2000) School performance and choice: The Chilean experience. *The Journal of Human Resources,* 35 (2), 392-417.

Mizala, A. and Romaguera, P. (2002) Equity and educational performance. Santiago: Universidad de Chile.

Mizala, A., Romaguera, P. and Ostoic, C. (2004) A hierarchical model for studying equity and achievement in the Chilean school choice system. Santiago: Universidad de Chile.

Mizala, A. and Torche, F. (2012) Bringing the schools back in: The stratification of achievement in the Chilean voucher system. *International Journal of Educational Development,* 32 (1), 132-144.

Moerbeek, M. (2004) The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research,* 39 (1), 129-149.

Mok, M. M. C., Kennedy, K. J. and Moore, P. J. (2011) Academic attribution of secondary students: Gender, year level and achievement level. *Educational Psychology,* 31 (1), 87-104.

Mok, M. M. C., Mcinerney, D. M., Zhu, J. and Or, A. (2015) Growth trajectories of mathematics achievement: Longitudinal tracking of student academic progress. *British Journal of Educational Psychology,* 85 (2), 154-171.

Monk, D. H. (1992) Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis,* 14 (4), 307-332.

Monk, D. H. (1994) Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review,* 13 (2), 125-145.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1988) *School matters: The junior years.* Wells: Open Books.

Muijs, D., Campbell, J., Kyriakides, L. and Robinson, W. (2005) Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement,* 16 (1), 51-70.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H. and Earl, L. (2014) State of the art: Teacher effectiveness and professional learning. *School Effectiveness and School Improvement,* 25 (2), 231-256.

Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L. and Stemler, S. E. (2000) *Gender differences in achievement, IEA's third international mathematics and science study.* Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P. and Arora, A. (2012) TIMSS 2011 international results in mathematics. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muñoz, M. A. and Chang, F. C. (2007) The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education,* 20 (3), 147-164.

Muñoz-Chereau, B. and Thomas, S. M. (2015) Educational effectiveness in Chilean secondary education: Comparing different 'value added' approaches to evaluate schools. *Assessment in Education: Principles, Policy & Practice*, 1-27.

Muñoz-Izquierdo, C. (1996) *Origen y consecuencias de las desigualdades educativas: Investigaciones realizadas en América Latina sobre el problema [Origin and consequences of educational inequalities: Research carried out in Latin America on this issue].* Mexico DF: Fondo de Cultura Económica.

Murillo, F. (2007a) *Investigación iberoamericana sobre eficacia escolar (IIEE) [Ibero-american research on school effectiveness (IIEE)].* Bogotá: Convenio Andrés Bello.

Murillo, F. (2007b) School effectiveness research in Latin America. *In:* T. Townsend (ed.) *International handbook of school effectiveness and improvement.* Netherlands: Springer.

Murillo, F. and Roman, M. (2011) The school or the cradle? Evidence of its contribution to students' performance in Latin America. Multilevel study about the estimate of school effects. *Profesorado,* 15, 26-50.

Muthén, B. O. and Khoo, S. T. (1998) Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences,* 10 (2), 73-101.

Muthén, L. K. and Muthén, B. O. (2010) *MPLUS user's guide.* Los Angeles, CA: Muthén & Muthén.

Nuttall, D., Goldstein, H., Prosser, R. and Rasbach, J. (1989) Differential school effectiveness. *International Journal of Educational Research,* 13, 769-776.

Nye, B., Konstantopoulos, S. and Hedges, L. V. (2004) How large are teacher effects? *Educational Evaluation and Policy Analysis,* 26 (3)**,** 237-257.

O'Donnell, S. and Sargent, C. (2011) International review of curriculum and assessment frameworks. INCA comparative tables. Available: http://www.inca.org.uk/ [Accessed 21 June 2012].

Oakes, J. (1985) *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

OECD (2009) *PISA 2006 technical report.* Paris: OECD Publishing.

OECD (2013a) *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed (volume II).* Paris: OECD Publishing.

OECD (2013b) *PISA 2012 results: What makes schools successful? Resources, policies and practices (volume IV).* Paris: OECD Publishing.

OECD (2014a) *PISA 2012 results: What students know and can do – student performance in mathematics, reading and science (Volume I, revised edition, February 2014).* Paris: OECD Publishing.

OECD (2014b) *TALIS 2013 results: An international perspective on teaching and learning.* Paris: OECD Publishing.

OECD (2015a) *Education at a glance 2015: OECD indicators.* Paris: OECD Publishing.

OECD (2015b) *Education policy outlook 2015: Making reforms happen.* Paris: OECD Publishing.

OECD/CAF/ECLAC (2014) Latin American economic outlook 2015: Education, skills and innovation for development. Paris: OECD Publishing.

Opdenakker, M., Van Damme, J., De Fraine, B., Van Landeghem, G. and Onghena, P. (2002) The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvement,* 13 (4)**,** 399-427.

Opdenakker, M.-C. and Van Damme, J. (2006) Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and catholic schools and types of schools. *School Effectiveness and School Improvement,* 17 (1)**,** 87-117.

Ortúzar, M., Flores, C., Milesi, C. and Cox, C. (2009) Aspectos de la formación inicial docente y su influencia en el rendimiento académico de los alumnos [Aspects of initial teacher treaning and its influence on student acedemic achievement]. *Camino al bicentenario, propuestas para Chile.* Santiago: Ediciones Universidad Católica.

Palardy, G. J. (2008) Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement,* 19 (1)**,** 21-49.

Palardy, G. J. (2010) The multilevel crossed random effects growth model for estimating teacher and school effects: Issues and extensions. *Educational and Psychological Measurement,* 70 (3)**,** 401-419.

Palardy, G. J. and Peng, L. (2015) The effect of summer on value-added assessments of teacher and school performance. *Education Policy Analysis Archives,* 23 (92)**,** 1-29.

Panel of Experts for Quality Education (2010) Informe final: Primera etapa. Propuestas para fortalecer la profesión docente en el sistema escolar chileno [Final report: First stage. Proposals to strengthen the teaching profession in the Chilean school system]. Santiago: Ministerio de Educación.

Peetsma, T., van der Veen, I., Koopman, P. and van Schooten, E. (2006) Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement,* 17 (3), 275-302.

Pianta, R. C., La Paro, K. and Hamre, B. K. (2008) *Classroom assessment scoring system.* Baltimore: Paul H. Brookes.

Pinxten, M., Marsh, H., De Fraine, B., Van Den Noortgate, B. and Van Damme, J. (2014) Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology,* 84 (1), 152-174.

Pustjens, H., Van de gaer, E., Van Damme, J., Onghena, P. and Van Landeghem, G. (2007) The short-term and the long-term effect of primary schools and classes on mathematics and language achievement scores. *British Educational Research Journal,* 33 (3), 419-440.

Raftery, A. E. and Lewis, S. M. (1992) Bayesian statistics 4. In*:* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.) *How many iterations in the gibbs sampler?* Oxford: Oxford University Press.

Ramírez, M. J. (2006) Factors related to mathematics achievement in Chile. In: S. J. H. T. Plomp (ed.) *Context of learning mathematics and science: Lessons learned from timss.* London: Routledge.

Rasbash, J. and Browne, W. J. (2001) Modelling non-hierarchical structures. In: A. H. Leyland and H. Goldstein (eds.) *Multilevel modelling of health statistics.* Chichester: John Wiley & Sons.

Rasbash, J., Leckie, G., Pillinger, R. and Jenkins, J. (2010) Children's educational progress: Partitioning family, school and area effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 173 (3), 657-682.

Raudenbush, S. W. (1989) The analysis of longitudinal, multilevel data. *International Journal of Educational Research,* 13 (7), 721-740.

Raudenbush, S. W. (1993) A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics,* 18 (4), 321-349.

Raudenbush, S. W. (1995) Hierarchical linear models to study the effects social context on development. *In:* J. M. Gottman (ed.) *The analysis of change.* Mahwah: Lawrence Earlbaum Associates.

Raudenbush, S. W. (2004) What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics,* 29 (1), 121-129.

Raudenbush, S. W. (2008) Many small groups. In: J. de Leeuw and E. Meijer (eds.) *Handbook of multilevel analysis.* New York: Springer.

Raudenbush, S. W. and Bryk, A. S. (2002) *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks: Sage.

Raudenbush, S. W. and Chan, W. (1993) Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting & Clinical Psychology,* 61 (6), 941-951.

Raudenbush, S. W. and Liu, X. (2000) Statistical power and optimal design for multisite randomized trials. *Psychological methods,* 5 (3), 199-213.

Raudenbush, S. W. and Willms, J. D. (1995) The estimation of school effects. *Journal of Educational and Behavioral Statistics,* 20 (4), 307-335.

Reezigt, G. J., Guldemond, H. and Creemers, B. P. M. (1999) Empirical validity for a comprehensive model on educational effectiveness. *School Effectiveness and School Improvement,* 10 (2)**,** 193-216.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. and Stringfield, S. (2014) Educational effectiveness research (eer): A state-of-the-art review. *School Effectiveness and School Improvement,* 25 (2)**,** 197-230.

Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005) Teachers, schools, and academic achievement. *Econometrica,* 73 (2)**,** 417-458.

Rockoff, J. E. (2004) The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review,* 94 (2)**,** 247-252.

Rodríguez Menés, J. and Donato, L. (2015) Social capital, social cohesion and cognitive attainment. In: Y. Li (ed.) *Handbook of research methods and applications in social capital.* Cheltenham: EdwardElgar.

Rogosa, D., Brand, D. and Zimowski, M. (1982) A growth curve approach to the measurement of change. *Psychological Bulletin,* 92 (3)**,** 726-748.

Rowan, B., Correnti, R. and Miller, R. J. (2002) What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *Teachers College Record,* 104 (8)**,** 1525-1567.

Rubin, D. B. (1987) *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rubin, D. B., Stuart, E. A. and Zanutto, E. L. (2004) A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics,* 29 (1)**,** 103-116.

Rumberger, R. W. (2005) Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record,* 107 (9)**,** 1999 -2045.

Rumberger, R. W. and Palardy, G. J. (2004) Multilevel models for school effectiveness research. In: D. Kaplan (ed.) *Handbook on quantitative methodology for the social sciences.* Thousand Oaks: Sage.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J. and Smith, A. (1979) *Fifteen thousand hours: Secondary schools and their effects on children.* London: Open Books.

Salganik, L. H. (1994) Apples and apples: Comparing performance indicators for places with similar demographic characteristics. *Educational Evaluation and Policy Analysis,* 16 (2)**,** 125-141.

Sammons, P., Day, C., Alison, K., Gu, Q., Gordon, S. and Smees, R. (2007) Exploring variations in teachers' work, lives and their effects on pupils: Key findings and implications from a longitudinal mixed-method study. *British Educational Research Journal,* 33 (5)**,** 681-701.

Sammons, P., Hillman, J. and Mortimore, P. (1995) *Key characteristics of effective schools: A review of school effectiveness research.* London: Office for Standards in Education and Institute of Education, University of London.

Sammons, P. and Luyten, H. (2009) Editorial article for special issue on alternative methods for assessing school effects and schooling effects. *School Effectiveness and School Improvement,* 20 (2)**,** 133-143.

Sammons, P., Melhuish, E., Siraj-Blatchford, I., Taggart, B. and Hunt, S. (2008) Effective pre-school and primary education 3-11 project (EPPE 3-11): Influences on children's attainment and progress in Key Stage 2: Cognitive outcomes in year 6. London: Department for Education, UK Government.

Sammons, P., Mortimore, P. and Thomas, S. (1996) Do schools perform consistently across outcomes and areas? In: D. Gray, D. Reynolds, C. Fitz-Gibbon and D. Jesson (eds.) *Merging traditions: The future of research on school effectiveness and school improvement.* London: Cassell.

Sammons, P., Nuttall, D. and Cuttance, P. (1993) Differential school effectiveness: Results from a reanalysis of the inner London education authority's junior school project data. *British Educational Research Journal,* 19 (4)**,** 381-405.

Sammons, P., Thomas, S. and Mortimore, P. (1997) *Forging links : Effective schools and effective departments.* London: Paul Chapman.

San Martín, E. and Carrasco, A. (2012) Clasificación de escuelas en la nueva institucionalidad educativa: Contribución de modelos de valor agregado para una responsabilización justa [Classification of schools in the new educational institutional environment: The contribution of value-added models to fair accountability]. *Temas de La Agenda Pública,* 7 (53)**,** 1-18.

Sanders, W. L. 2006. Comparisons among various educational assessment value-added models. *Power of Two-National Value-Added Conference.* Columbus.

Sanders, W. L. and Rivers, J. C. (1996) Cumulative and residual effects of teachers on future student academic achievement. Knoxville: University of Tennessee.

Sanders, W. L., Wright, S. P. and Horn, S. P. (1997) Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education,* 11**,** 57-67.

Santiago, P., Benavides, F., Danielson, C., Goe, L. and Nusche, D. (2013) *Teacher evaluation in Chile 2013, OECD reviews of evaluation and assessment in education.* Paris: OECD Publishing.

Sapelli, C. and Vial, B. (2002) The performance of private and public schools in the Chilean voucher system. *Cuadernos de Economía,* 39 (118)**,** 423-454.

Schafer, J. L. (1997) *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Schafer, J. L. and Graham, J. W. (2002) Missing data: Our view of the state of the art. *Psychological methods,* 7 (2)**,** 147-177.

Schagen, I. and Hutchison, D. (2003) Adding value in educational research. The marriage of data and analytical power. *British Educational Research Journal,* 29**,** 749–765.

Scheerens, J. (1999) School effectiveness in developed and developing countries: A review of the research evidence (Human Development Network Report). Washington, DC: The World Bank.

Scheerens, J. (2000) *Improving school effectiveness.* Paris: UNESCO.

Scheerens, J. (2001) Monitoring school effectiveness in developing countries. *School Effectiveness and School Improvement,* 12 (4)**,** 359-384.

Scheerens, J. and Bosker, R. J. (1997) *The foundations of educational effectiveness.* New York: Pergamon.

Seidel, T. and Shavelson, R. J. (2007) Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research,* 77 (4)**,** 454-499.

Seltzer, M. H., Wong, W. H. and Bryk, A. S. (1996) Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics,* 21 (2)**,** 131-167.

Singer, J. and Willett, J. B. (2003) *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford: Oxford University Press.

Sirin, S. R. (2005) Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research,* 75 (3), 417-453.

Snijders, T. A. B. and Bosker, R. J. (2012) *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. v. d. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology),* 64 (4), 583-639.

Strand, S. (2010) Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement,* 21 (3), 289-314.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2010) *Early childhood matters: Evidence from the effective pre-school and primary education project.* London: Routledge.

Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., Bankow, K., Rodriguez, M. and Reckase, M. D. (2012) *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA teacher education and development study in mathematics (TEDS-M).* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D. and Yu, F. (2006) The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation,* 12 (6), 561-582.

Teddlie, C. and Reynolds, D. (eds.) (2000) *The international handbook of school effectiveness research,* London: Falmer.

Televantou, I., Marsh, H., Kyriakides, L., Nagengast, B., Fletcher, J. and Malmberg, L. (2015) Phantom effects in school composition research: Consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement,* 26 (1), 75-101.

Thomas, S. (2001) Dimensions of secondary school effectiveness: Comparative analyses across regions. *School Effectiveness and School Improvement,* 12 (3), 285-322.

Thomas, S., Sammons, P., Mortimore, P. and Smees, R. (1997a) Differential secondary school effectiveness: Comparing the performance of different pupil groups. *British Educational Research Journal,* 23 (4), 451-469.

Thomas, S., Sammons, P., Mortimore, P. and Smees, R. (1997b) Stability and consistency in secondary schools' effects on students' GCSE outcomes over three years. *School Effectiveness and School Improvement,* 8 (2), 169-197.

Thrupp, M. (1999) *Schools making a difference: Let's be realistic!* Buckingham: Open University Press.

Thum, Y. M. (2003) Measuring progress toward a goal. Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods & Research,* 32 (2), 153-207.

Timmermans, A. C., Doolaard, S. and de Wolf, I. (2011) Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness and School Improvement,* 22 (4), 393-413.

Timmermans, A. C., Snijders, T. A. B. and Bosker, R. J. (2013) In search of value added in the case of complex school effects. *Educational and Psychological Measurement,* 73 (2), 210-228.

Timmermans, A. C. and Thomas, S. M. (2015) The impact of student composition on schools' value-added performance: A comparison of seven empirical studies. *School Effectiveness and School Improvement,* 26 (3), 487-498.

Tokman, A. (2002) Is private education better? Evidence from Chile. Santiago: Central Bank of Chile.

Tonry, M. H., Ohlin, L. E., Farrington, D. P., Development, P. o. H. and Behavior, C. (1991) *Human development and criminal behavior: New ways of advancing knowledge.* New York: Springer.

Townsend, T. (ed.) (2007) *International handbook of school effectiveness and improvement.* Netherlands: Springer.

Treviño, E. and Donoso, F. (2010) Agrupación de escuelas para intervenciones de política: Análisis del caso chileno [Classification of schools for policy interventions: Analysis of the Chilean case]. Santiago: Universidad Diego Portales.

Treviño, E., Toledo, G. and Gempp, R. (2013) Preschool education quality: Teacher practices and the path to improvement. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana,* 50 (1), 40-62.

Treviño, E., Valdés, H., Castro, M., Costilla, R., Pardo, C. and Rivas, F. D. (2010) Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe [Factors associated with cognitive outcomes of students in Latin America and the Caribbean]. Santiago: UNESCO.

Troncoso, P., Pampaka, M. and Olsen, W. (2015) Beyond traditional school value-added models: A multilevel analysis of complex school effects in Chile. *School Effectiveness and School Improvement* [Online].

Tymms, P., Merrell, C. and Henderson, B. (2000) Baseline assessment and progress during the first three years at school. *Educational Research and Evaluation,* 6 (2), 105-129.

UNESCO (2008a) *Eficacia escolar y factores asociados en América Latina y el Caribe [School effectiveness and factors associated in Latin America and the Caribbean].* Santiago: UNESCO-LLECE.

UNESCO (2008b) Student achievement in Latin America and the Caribbean. Results of the second regional comparative and explanatory study (SERCE). Santiago: Regional Bureau for Education in Latin America Latina and the Caribbean OREALC/UNESCO.

UNESCO 2014. Education for all (EFA) in Latin America and the Caribbean: Assessment of progress and post-2015 challenges. Lima: UNESCO.

UNESCO (2015a) Latin America and the caribbean: Regional overview. Paris: UNESCO.

UNESCO (2015b) *Tercer estudio regional comparativo y explicativo [Third regional comparative and explicative study]. Factores asociados [Associated factors].* Paris: UNESCO.

UNESCO (2015c) *Tercer estudio regional comparativo y explicativo [Third regional comparative and explicative study]. Logros de aprendizajes [Learning achievement].* Paris: UNESCO.

United Nations Development Programme (2015) Human development report 2015. Briefing note for countries on the 2015 human development report: Chile. New York: United Nations Development Programme.

Valenzuela, J. P., Bellei, C. and de los Ríos, D. (2014) Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy,* 29 (2)**,** 217-241.

van Buuren, S. (2011) Multiple imputation of multilevel data. *In:* J. Hox and J. K. Roberts (eds.) *Handbook of advanced multilevel analysis.* New York: Routledge.

van de gaer, E., De Fraine, B., Pustjens, H., van Damme, J., De Munter, A. and Onghena, P. (2009) School effects on the development of motivation toward learning tasks and the development of academic self-concept in secondary education: A multivariate latent growth curve approach. *School Effectiveness and School Improvement,* 20 (2)**,** 235-253.

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,* 16 (3)**,** 281-303.

van Ewijk, R. and Sleegers, P. (2010a) The effect of peer socioeconomic status on achievement: A meta-analysis. *Educational Research Review,* 5 (2)**,** 134-150.

van Ewijk, R. and Sleegers, P. (2010b) Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement,* 21 (3)**,** 237-265.

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,* 40 (4)**,** 423-434.

Vasquez-Salgado, Y. and Chavira, G. (2014) The transition from middle school to high school as a developmental process among latino youth. *Hispanic Journal of Behavioral Sciences,* 36 (1)**,** 79-94.

Velez, E., Schiefelbein, E. and Valenzuela, J. (1993) Factors affecting achievement in primary education. Washington: The World Bank.

Verhaeghe, J. P., Van Damme, J. and Knipprath, H. 2011. *Value added of primary schools with high proportions of minority students: A longitudinal study*. Paper presented at the International Congress for School Effectivenes and Improvement (ICSEI), Limassol, Cyprus, January. Unpublished.

Walsh, J. E. (1947) Concerning the effect of intraclass correlation on certain significance tests. *The Annals of Mathematical Statistics,* 18 (1)**,** 88-96.

Wayne, A. J. and Youngs, P. (2003) Teacher characteristics and student achievement gains: A review. *Review of Educational Research,* 73 (1)**,** 89-122.

White, K. R. (1982) The relation between socioeconomic status and academic achievement. *Psychological Bulletin,* 91 (3)**,** 461-481.

Wilkins, J. L. M. and Ma, X. (2002) Predicting student growth in mathematical content knowledge. *The Journal of Educational Research,* 95 (5)**,** 288-298.

Willett, J. B. (1988) Questions and answers in the measurement of change. *Review of Research in Education,* 15**,** 345-422.

Willms, J. D. (1985) The balance thesis: Contextual effects of ability on pupils' O-grade examination results. *Oxford Review of Education,* 11 (1)**,** 33-41.

Willms, J. D. (1992) *Monitoring school performance: A guide for educators.* London: Falmer Press.

Willms, J. D. (2002) *Vulnerable children: Findings from Canada's national longitudinal survey of children and youth.* Edmonton: University of Alberta Press.

Willms, J. D. (2010) School composition and contextual effects on student outcomes. *Teachers College Record,* 112 (4), 1008-1037.
Willms, J. D. and Jacobsen, S. (1990) Growth in mathematics skills during the intermediate years: Gender differences and school effects. *International Journal of Educational Research,* 14, 157-174.
Willms, J. D. and Raudenbush, S. W. (1989) A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement,* 26 (3), 209-232.
Willms, J. D. and Somer, M.-A. (2001) Family, classroom, and school effects on childrens educational outcomes in Latin America. *School Effectiveness and School Improvement,* 12 (4), 409-445.
Wilson, S., Floden, R. and Ferrini-Mundy, J. (2001) Teacher preparation research: Current knowledge, gaps, and recommendations. Washington: University of Washington.
Winkler, W. (1995) Matching and record linkage. *In:* B. Cox (ed.) *Business survey methods.* New York: Wiley.
World Bank (2015) *World development indicators: Gini index.* Washington, DC: World Bank.
Yang, Y. and Gustafsson, J. E. (2004) Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation,* 10 (3), 259-288.
Zvoch, K. and Stevens, J. J. (2003) A multilevel, longitudinal analysis of middle school math and language achievement. *Education Policy Analysis Archives,* 11 (20), 1-21.
Zvoch, K. and Stevens, J. J. (2006) Successive student cohorts and longitudinal growth models: An investigation of elementary school *Education Policy Analysis Archives,* 14 (2), 1-25.

APPENDICES

APPENDIX 1: DESCRIPTION OF SOURCES OF SECONDARY DATA

The main features of the assessment projects and official systems of information from which secondary data was retrieved are described in the following section.

The SEPA Project

The SEPA[41] Project is a student assessment initiative implemented by the MIDE UC Assessment Centre of the Pontificia Universidad Católica de Chile[42] since 2007. The main purpose of this assessment programme is to report to individual schools on the academic progress their students make each year, in comparison to other students attending similar schools, with the aim of informing pedagogical decisions (MIDE UC, 2015). The project assesses student achievement progress in the primary and secondary school levels by collecting longitudinal data on individual student achievement in language (Spanish) and mathematics. The SEPA tests are designed to measure student performance on the grade-level competencies specified in the national curriculum standards and, therefore, are criterion-referenced assessments. The test scores in language and mathematics are both vertically and horizontally scaled by statisticians at the MIDE UC Centre, using item response theory (IRT). This makes it possible to, within each subject, compare scores among the different grades assessed and between the yearly assessment cycles.

Thus, the SEPA Project offers to its clients (individual schools, school consortiums and municipal corporations that administrate groups of public schools) a compared longitudinal analysis of students' growth in academic achievement. The following table shows the number of students and schools that have participated in the project since in place and until 2012.

[41] SEPA is an acronym standing for Learning Progress Evaluation System. Further information available on the website www.sepauc.cl.

[42] The MIDE UC Assessment Centre of the Pontificia Universidad Católica de Chile is a non-profit research, development, and service organisation that conducts measurement and evaluation applied to various fields, such as education, organizations and society, among others. For years, the Ministry of Education has commissioned the MIDE UC Centre to develop several assessment initiatives. The researcher worked in the Centre as a data analyst and research coordinator prior to the beginning of her doctoral studies, and has, therefore, a deep insight into the content and structure of the data collected by the organisation.

TABLE 25: NUMBER OF STUDENTS ASSESSED WITH THE SEPA LANGUAGE TESTS BY YEAR AND GRADE

LANGUAGE							
	2007	**2008**	**2009**	**2010**	**2011**	**2012**	**Total**
1 Primary School	1,180	1,662	2,416	4,521	4,435	4,415	18,629
2	544	2,471	2,838	5,853	5,068	5,066	21,840
3	1,820	3,630	2,828	7,019	5,923	5,758	26,978
4	215	2,332	3,816	5,223	4,410	4,730	20,726
5	1,103	5,086	3,081	4,020	4,843	4,756	22,889
6	571	2,169	5,695	5,200	4,195	4,559	22,389
7	1,076	5,353	2,435	6,318	4,820	5,519	25,521
8	-	2,571	10,961	4,977	3,711	4,265	26,485
1 Secondary School	-	-	3,571	8,466	7,263	4,155	23,455
2	-	-	-	3,049	4,414	3,388	10,851
3	-	-	-	-	2,200	2,235	4,435
Total	**6,509**	**25,274**	**37,641**	**54,646**	**51,282**	**48,846**	**224,198**

TABLE 26: NUMBER OF STUDENTS ASSESSED WITH THE SEPA MATHEMATICS TESTS BY YEAR AND GRADE

MATHEMATICS							
	2007	**2008**	**2009**	**2010**	**2011**	**2012**	**Total**
1 Primary School	1,224	1,501	2,086	3,574	4,159	4,166	**16,710**
2	562	2,370	2,599	5,361	4,591	4,992	**20,475**
3	1,834	3,511	2,652	6,321	5,555	5,600	**25,473**
4	217	2,209	3,632	4,652	4,387	4,602	**19,699**
5	1,119	4,963	2,924	3,895	4,795	4,660	**22,356**
6	576	2,049	5,624	5,030	4,072	4,480	**21,831**
7	1,091	5,333	2,332	6,059	4,642	5,440	**24,897**
8	-	2,554	11,087	5,451	4,410	4,184	**27,686**
1 Secondary School	-	-	3,627	8,290	8,455	4,844	**25,216**
2	-	-	-	2,938	5,032	4,141	**12,111**
3	-	-	-	-	2,191	2,270	**4,461**
Total	**6,623**	**24,490**	**36,563**	**51,571**	**52,289**	**49,379**	**220,915**

TABLE 27: NUMBER OF SCHOOLS PARTICIPATING IN THE SEPA PROJECT BY SUBJECT AND YEAR

	LANGUAGE	MATHEMATICS
2007	37	37
2008	121	120
2009	203	200
2010	198	186
2011	157	157
2012	181	181
Total	**897**	**881**

The SEPA project's set of instruments consists of multiple-choice items tests, ranging from 20 items in 1st grade tests to 50 items in 8th grade and higher tests. The following table shows the specific number of items, average difficulty and reliability for the tests administered by the project between 2010 and 2012 to the cohorts analysed in the present study. The high Cronbach's α coefficients, ranging from .847 to .928, indicate that the SEPA tests are internally homogeneous.

TABLE 28: NUMBER OF ITEMS, RELIABILITY (CRONBACH'S ALPHA) AND AVERAGE DIFFICULTY OF SEPA TESTS BY SUBJECT, COHORT, GRADE LEVEL AND FORM

	Grade Level		3		4		5		6		7		8	
	Form		**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**
LANGUAGE	**Cohort I**	N items	35	35	40	40	40		-	-	-	-	-	-
		Reliability	0.902	0.894	0.864	0.876	0.885		-	-	-	-	-	-
		Difficulty	0.421	0.401	0.428	0.429	0.563		-	-	-	-	-	-
	Cohort II	N items	-	-	40	40	40	40	40		-	-	-	-
		Reliability	-	-	0.911	0.908	0.887	0.882	0.864		-	-	-	-
		Difficulty	-	-	0.407	0.402	0.513	0.519	0.55		-	-	-	-
	Cohort III	N items	-	-	-	-	40	40	40	40	50		-	-
		Reliability	-	-	-	-	0.917	0.917	0.867	0.875	0.899		-	-
		Difficulty	-	-	-	-	0.395	0.383	0.521	0.533	0.560		-	-
	Cohort IV	N items	-	-	-	-	-	-	40	40	50	50	50	
		Reliability	-	-	-	-	-	-	0.914	0.917	0.900	0.901	0.884	
		Difficulty	-	-	-	-	-	-	0.427	0.424	.0503	0.497	0.560	
MATHEMATICS	**Cohort I**	N items	35	35	40	40	40		-	-	-	-	-	-
		Reliability	0.919	0.919	0.857	0.847	0.871		-	-	-	-	-	-
		Difficulty	0.467	0.47	0.551	0.551	0.55		-	-	-	-	-	-
	Cohort II	N items	-	-	40	40	40	40	40		-	-	-	-
		Reliability	-	-	0.909	0.903	0.856	0.868	0.883		-	-	-	-
		Difficulty	-	-	0.407	0.395	0.521	0.522	0.531		-	-	-	-
	Cohort III	N items	-	-	-	-	40	40	40	40	50		-	-
		Reliability	-	-	-	-	0.928	0.927	0.884	0.889	0.882		-	-
		Difficulty	-	-	-	-	0.407	0.398	0.534	0.542	0.527		-	-
	Cohort IV	N items	-	-	-	-	-	-	40	40	50	50	50	
		Reliability	-	-	-	-	-	-	0.923	0.922	0.879	0.885	0.871	
		Difficulty	-	-	-	-	-	-	0.412	0.409	0.487	0.478	0.485	

Note: In 2012 only one form of each test was administered. Dashes indicate that data are not available in that grade level for that cohort.

The SEPA project team has selected and released to the public some of the items used in previous test administrations. Four of these items, differing on their average difficulty and targeted grade, are presented below.

The following item was part of the 2010 Language test for primary students in Year 3. It assesses the 'reading' dimension, corresponds to the indicator "to identify a relationship of explicit facts in a non-literary text" and it was classified as a medium-difficulty item.

FIGURE 47: RELEASED ITEM N. 1 - 2010 LANGUAGE TEST FOR YEAR 3 STUDENTS

Los castores son roedores semi-acuáticos nativos de América del Norte y Europa. Son de color café, excepto su cola que es negra. Las hembras tienen entre 3 y 4 crías.

Son grandes arquitectos: talan árboles y embalsan las corrientes de agua para hacer lagos donde se ponen a salvo. Los diques que ellos forman llegan a medir más de 500 m de largo y son tan resistentes que soportan el peso de una persona. Los castores están adaptados a la vida en el agua, ya que tiene patas palmeadas y cola aplanada.

Se alimentan de corteza de árboles y hojas, y almacenan ramas bajo el agua para el invierno. Se alojan en un enorme montículo de ramas, que construyen en el centro del lago. Las entradas se encuentran bajo el agua, de modo que pueden entrar y salir sin ser vistos.

http://es.wikipedia.org

¿Qué características les han permitido a los castores adaptarse para vivir en el agua?

A Sus firmes pulmones y su fuerza.

B Sus largas aletas y su gran tamaño.

C Su piel gruesa y su resistencia al frío.

D Sus patas palmeadas y su cola aplanada. ✓

Also assessing language, the following item was part of the 2010 test for students in Year 8 of primary school. The item corresponds to the dimension 'communication and media', refers to the indicator "to distinguish facts from opinions in different types of texts" and is a high-difficulty item.

FIGURE 48: RELEASED ITEM N. 2 - 2010 LANGUAGE TEST FOR YEAR 8 STUDENTS

Lee y responde la siguiente pregunta:

> En un éxito total se ha transformado el grupo musical llamado "Daniel Muñoz, Félix Llancafil y 3x7 Veintiuna", que busca revitalizar el folclor nacional, mostrando una versión diferente de lo que puede ser la cueca. El primer disco del grupo, "Tricolores", salió en 2005 y tuvo una gran acogida. Por lo tanto, este año la agrupación sacará su segundo disco llamado "La otra patita".
>
> Esta fiebre "cuequera" se ha hecho sumamente popular en todo el país, por lo que el grupo está copado de presentaciones a pocos días de celebrar un nuevo aniversario patrio.
>
> Diario *Publimetro*

¿Cuál de las siguientes expresiones del texto anterior está libre de apreciaciones personales?

A "...busca revitalizar el folclor nacional, mostrando una versión diferente...".

B "En un éxito total se ha transformado el grupo musical...".

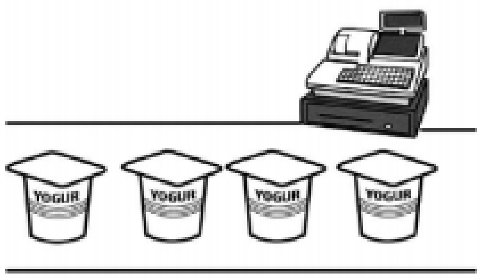
C "Esta fiebre "cuequera" se ha hecho sumamente popular...".

D "El primer disco del grupo, "Tricolores", salió en 2005...". ✓

In the 2010 mathematics test, the following item was targeted for students in Year 3. The dimension assessed is 'geometry' and the indicator name is "to solve problems that involve to determine a product in daily life contexts". The difficulty of this item is low.

FIGURE 49: RELEASED ITEM N. 3 - 2010 MATHEMATICS TEST FOR YEAR 3 STUDENTS

Una cajera pasa por caja los siguientes productos:



Si cada yogur vale $ 150, ¿cuál es el valor total de la venta?

A $ 400

B $ 420

C $ 600 ✓

D $ 604

Finally, also assessing mathematics, the item in the next figure was developed for Year 7 students, corresponds to the dimension 'numbers' and to the indicator "to solve a problem that requires to obtain the total number of elements, knowing the percentage that a certain quantity of them represents". In terms of difficulty, this item was classified as high.

FIGURE 50: RELEASED ITEM N. 4 - 2010 MATHEMATICS TEST FOR YEAR 7 STUDENTS

En el pueblo El Álamo, el 20 % de los habitantes son niños. Se sabe que en ese pueblo hay 3 500 niños. ¿Cuántos habitantes hay en El Álamo?

A 700

B 5 250

C 14 000

D 17 500 ✓

The School Enrolment Recording System

Since 1990, at the beginning of each academic year, the Ministry of Education conducts an official survey of the students enrolled in the schooling system. These public databases provide accurate information regarding the number of students for each school and classroom in the country.

General System of Student Information - SIGE

Every year since 2003, the General System of Student Information (SIGE) develops an administrative database with the list of teachers in the national system by school, grade, class and subject taught. The information is completed by schools' administrative staff via an on-line platform that is then validated by the Department of Studies of the Ministry of Education. Teachers are given a unique identification number that allows researchers to track them across time. These data permits the link between the information on teachers and students available in the other secondary sources. The number of cases available in the SIGE databases mirrors the number of teachers in the system as this registration system targets the population of teachers working in all Chilean schools and is the basis for official statistics.

The System for Measuring Educational Quality - SIMCE

The Chilean government has been systematically gathering data on student performance since 1988, with the aim of monitoring the quality of the educational system. This large-scale assessment programme, known as the System for Measuring Educational Quality (SIMCE)[43], tests all the students in certain grade levels on an annual basis, alternating between fourth grade, eighth grade and the second year of secondary school. Its design and administration is the responsibility of the Ministry of Education's Curriculum and Evaluation Unit and its main purpose is to generate reliable indicators on teaching quality (MINEDUC, 2012). The evaluation procedure is the same for every school in the country and is managed by external contractors. The scores on SIMCE tests reveal performance in language and mathematics (and, in some years, sciences). Tests started to include open-ended questions and introduced IRT methodology in 2000, allowing comparisons across years and making it possible to produce more accurate descriptions of different levels of performance, to measure with precision students with different skill levels, and to examine potential item bias.

These public databases contain reliable information regarding the characteristics of each school in the country. The SIMCE variables at the school level considered in the present study are schools' type of administration (municipal, private subsidised or private non-subsidised), geographical location (urban or rural), socio-economic level and overall mean achievement and variation in the SIMCE tests. At the student level, the parents of students assessed are asked to complete a questionnaire with information such as their level of education, family income and educational resources available at home, among other topics. These data on students and schools has been linked to the SEPA databases.

[43] Further information available on the website www.simce.cl.

This section deals with the analysis of missing data for the present study. Firstly, the configuration and location of observed and missing values in the data sets, this is, the patterns of missing data, are described. I then move on to the question of why the data are missing by analysing the missing data mechanisms. The analysis of missing data in this section does not include the data that are missing-by-design (i.e., the data of those students in grades that are not part of the considered grades in the accelerated longitudinal design). Both types of analyses were conducted using the SPSS Missing Values add-on module (IBM, 2011).

Missing Data Patterns

In what follows patterns of data missingness are analysed as a first step towards determining whether multiple imputation is necessary. These analyses refer to the location of missing values, how extensive they are and whether group of variables tend to have values missing in multiple cases. In Chapter 3, the methodological decisions with regard to the selection of data and the sample sizes at each step leading to the final list-wise deleted samples were presented. The analysis of missing data is carried out for the samples for which it would make sense to later impute missing values, this is the students that participated in the SEPA project, with plausible grade trajectories, belonging to one of the four cohorts of interest and that, additionally, were linked to a teacher for each of the three years of interest (i.e., 2010, 2011 and 2012) in both subjects. This reference stage was chosen, as it would not be reasonable to impute information about the teacher that taught the student when this information is not available. This gives a total of 20,373 and 19,492 students for language and mathematics, respectively.

In Figure 51 the numbers of missing values by variable are summarised in histograms for the language and the mathematics databases. These charts show that, on average, variables have missing values on 12 to 13% of the cases (2,634 in language and 2,433 in mathematics). Figure 52 shows histograms for the numbers of missing valued by cases. The figure indicate that cases in the language and mathematics databases have, on average, missing values on 12 to 13% of the 29 variables considered in the analysis (3.75 in language and 3.62 in mathematics).

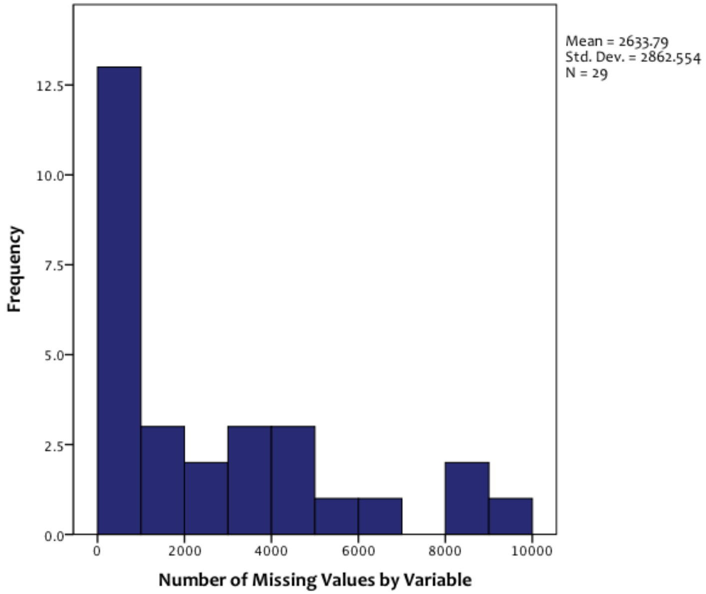
Table 29 provides a first look, variable by variable, at the extent of missing data. Here variables are sorted according to the level of analysis to which they refer, namely student-, teacher- and school-level variables.

The proportions of missing data by variable in the language data set are fairly similar to those in the mathematics data set. At the student-level, the variables *Number of books at home*, *Father's educational level*, *Mother's educational level* and *Home monthly income* have the greatest numbers of cases with missing values (between 19.0% and 21.7%). At the teacher-level, the variables *ITT programme duration* and *Major in the subject assessed* have the greatest numbers of cases with missing values, but they vary by year. The 2012 data for teachers have the greater proportion of missing values, followed by teachers in 2010. At the school level, the data is, to a large extent, complete. The completeness of school-level variables facilitates missing data imputation processes.

Figures 53 and 54 display three pie charts each that show different aspects of missing values in the data, when considering only the dependent variables, SEPA test scores in 2010, 2011 and 2012.

FIGURE 51: DISTRIBUTION OF NUMBER OF MISSING VALUES BY VARIABLE

LANGUAGE



MATHEMATICS

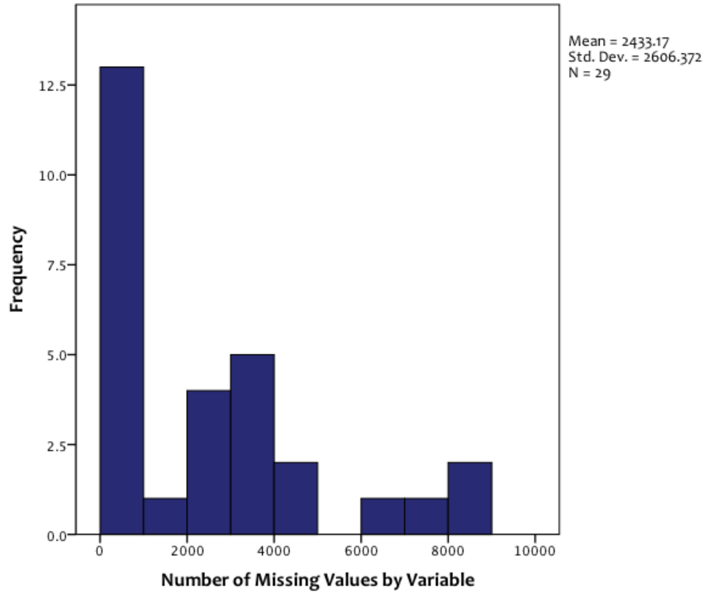
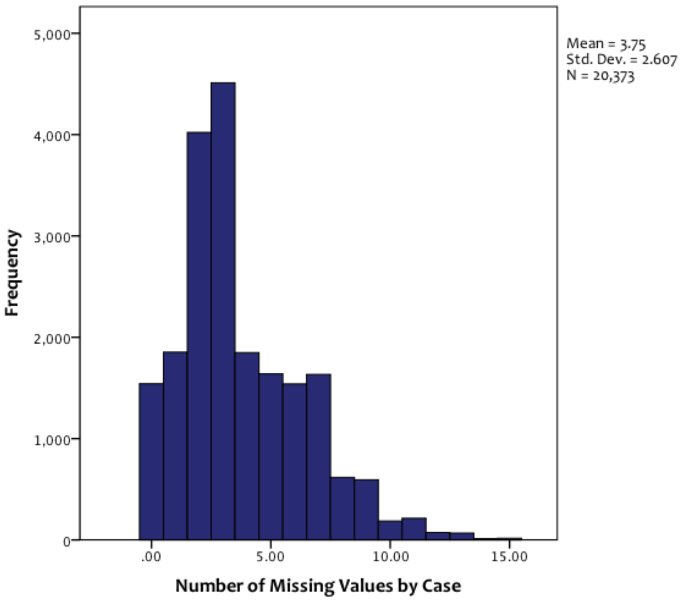


FIGURE 52: DISTRIBUTION OF NUMBER OF MISSING VALUES BY CASE

LANGUAGE



MATHEMATICS

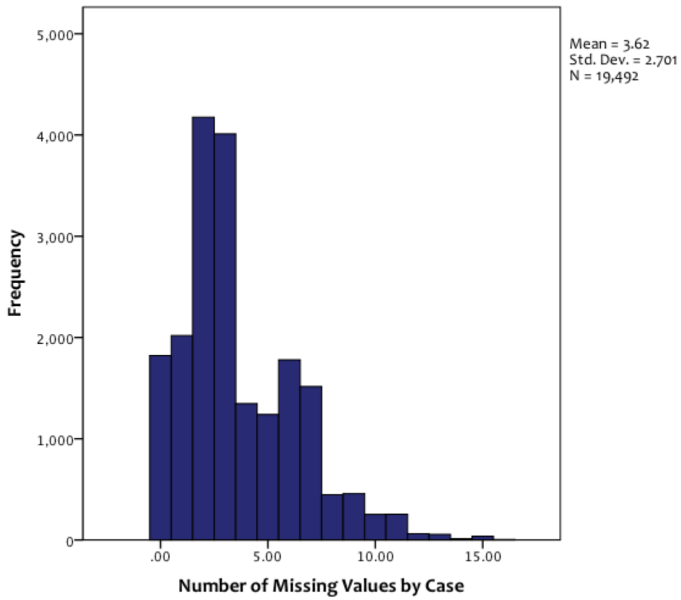
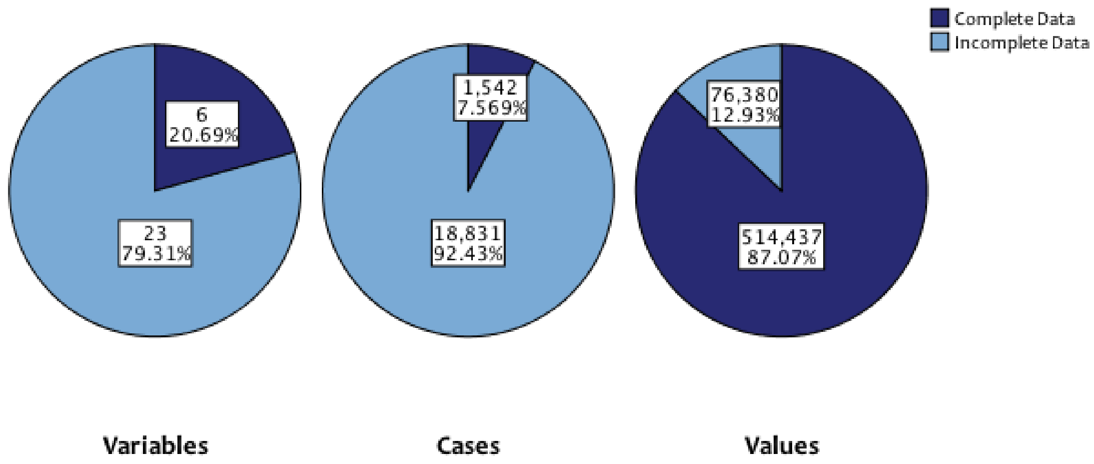


TABLE 29: MISSING DATA BY VARIABLE IN THE LANGUAGE AND MATHEMATICS DATA

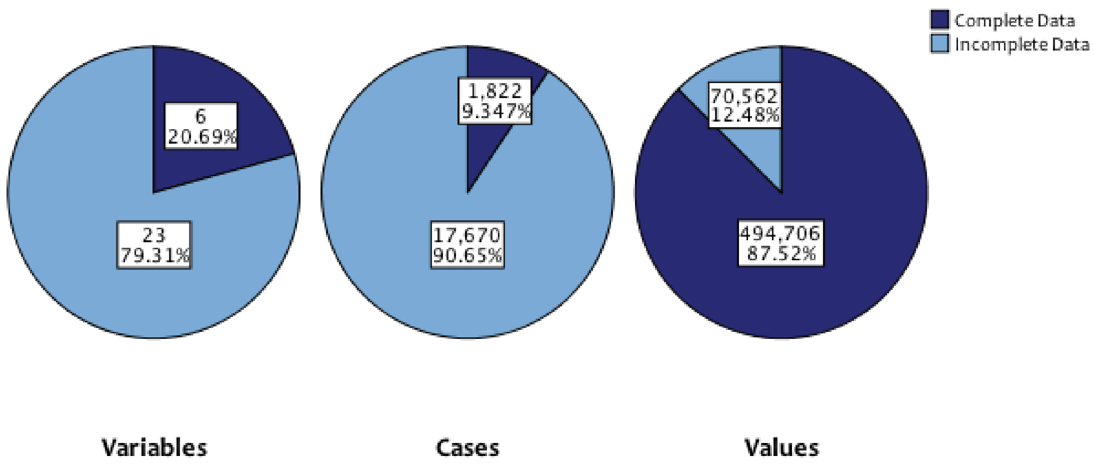
	Language			Mathematics		
	Valid N	Missing		Valid N	Missing	
		N	%		N	%
Within student variables						
SEPA score 2010	14,041	6,332	31.1	13,123	6,369	32.7
SEPA score 2011	11,399	8,974	44.0	11,303	8,189	42.0
SEPA score 2012	11,204	9,169	45.0	11,260	8,232	42.2
Student-level variables						
Gender	20,373	0	0.0	19,492	0	0.0
Age	20,373	0	0.0	19,492	0	0.0
Number of Books at Home	16,443	3,930	19.3	15,678	3,814	19.6
Mother's Educational Level	16,495	3,878	19.0	15,731	3,761	19.3
Father's Educational Level	15,958	4,415	21.7	15,214	4,278	21.9
Home Monthly Income	16,501	3,872	19.0	15,749	3,743	19.2
Teacher-level variables						
Gender 2010	18,504	1,869	9.2	17,435	2,057	10.6
Gender 2011	20,278	95	0.5	19,456	36	0.2
Gender 2012	19,560	813	4.0	18,852	640	3.3
Years of Experience 2010	18,504	1,869	9.2	17,435	2,057	10.6
Years of Experience 2011	20,278	95	0.5	19,456	36	0.2
Years of Experience 2012	19,560	813	4.0	18,852	640	3.3
ITT programme duration 2010	17,605	2,768	13.6	16,646	2,846	14.6
ITT programme duration 2011	19,034	1,339	6.6	18,227	1,265	6.5
ITT programme duration 2012	15,347	5,026	24.7	15,869	3,623	18.6
Major in the subject 2010	17,637	2,736	13.4	16,678	2,814	14.4
Major in the subject 2011	19,424	949	4.7	18,520	972	5.0
Major in the subject 2012	15,532	4,841	23.8	16,005	3,487	17.9
School-level variables						
Type (Dummy Private Non-Subsidised/Public)	12,206	8,167	40.1	12,202	7,290	37.4
Type (Dummy Private Subsidised/Public)	15,971	4,402	21.6	15,107	4,385	22.5
School SES	20,373	0	0.0	19,492	0	0.0
Rural	20,373	0	0.0	19,492	0	0.0
School Size	20,373	0	0.0	19,492	0	0.0
Student/Teacher Ratio	20,373	0	0.0	19,492	0	0.0
School Overall Performance	20,352	21	0.1	19,471	21	0.1
School SD Performance	20,366	7	0.0	19,485	7	0.0

FIGURE 53: OVERALL SUMMARY OF MISSING VALUES IN SEPA LANGUAGE TEST SCORES



Note: The variables chart shows in light blue the number and percentage of analysis variables have at least one missing value on a case. The cases chart shows in light blue number and percentage of the cases that have at least one missing value on a variable. The values chart shows the number and percentage of values (cases × variables) that are missing.

FIGURE 54: OVERALL SUMMARY OF MISSING VALUES IN SEPA MATHEMATICS TEST SCORES



Note: The variables chart shows in light blue the number and percentage of analysis variables have at least one missing value on a case. The cases chart shows in light blue number and percentage of the cases that have at least one missing value on a variable. The values chart shows the number and percentage of values (cases × variables) that are missing.

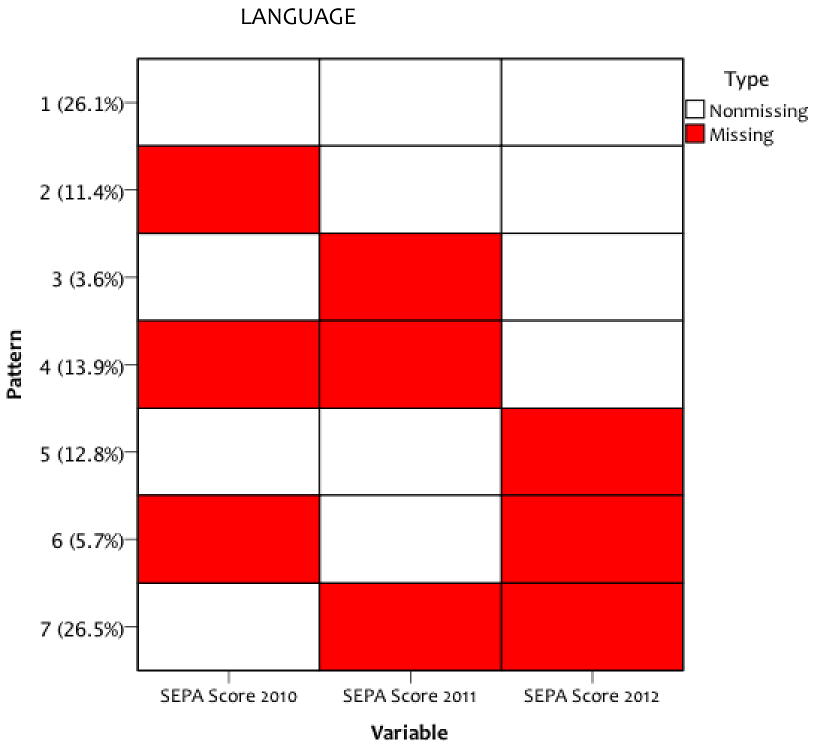
Unsurprisingly, 23 out of the 29 variables have at least one missing value on a case and more than 90% of the cases have at least one missing value on a variable. In the language data set, of the 590,817 cells (20,373 cases X 29 variables) in the data matrix, 76,380 of the cells (12.93%) were empty; that is, they had missing values. Similarly, in the mathematics dataset, of the 565,268 cells (19,492 cases X 29 variables) in the data matrix, 70,562 of the cells (12.48%) were empty. This confirms that, if listwise deletion were to be used, much of the information in the data sets would be lost.

Figure 55, in turn, corresponds to patterns charts displaying missing value patterns for the dependent variables in the language and mathematics databases. Here, each pattern corresponds to a group of cases with the same pattern of incomplete and complete data. There are seven patterns of jointly missing data that occur in more than 1% of the cases. For example, in both subjects, Pattern 1 represents cases that have no missing values, while Pattern 7 represents cases that have missing values on both SEPA Test score 2011 and SEPA Test score 2012. This figure also shows the percentage of cases under each pattern. Patterns 1 and 7 are the two most common patters, covering approximately 50% of the cases. The least common pattern is to have test scores only in 2011 (Pattern 3).

Tables 30 and 31 show whether the student-level data tend to be missing for multiple variables in individual cases in order to determine whether the data are jointly missing. The variables 'Home monthly income', Mother's educational level', 'Number of books at home' and 'Father's educational level' are very frequently jointly missing. This is not surprising as they all come from the same secondary source (the SIMCE Assessment System databases). The other student variables (i.e., gender and age) come from the Student Enrolment Recording System and are complete.Tables 32 and 33 show the missing patterns for the teacher-level variables. As expected, teacher variables are missing in groups according to year. Also, within years, data are missing jointly in two groups: on one hand gender and years of teaching experience and, on the other, information on their initial teacher training[44]. Finally, school variables, in both the language and mathematics data sets, are jointly complete in more than 99% of the cases.

[44] This might be due to the Teacher Census' strategy for collecting teacher data.

FIGURE 55: MISSING DATA PATTERNS IN SEPA TEST SCORES AND PERCENTAGE OF CASES FOR EACH PATTERN



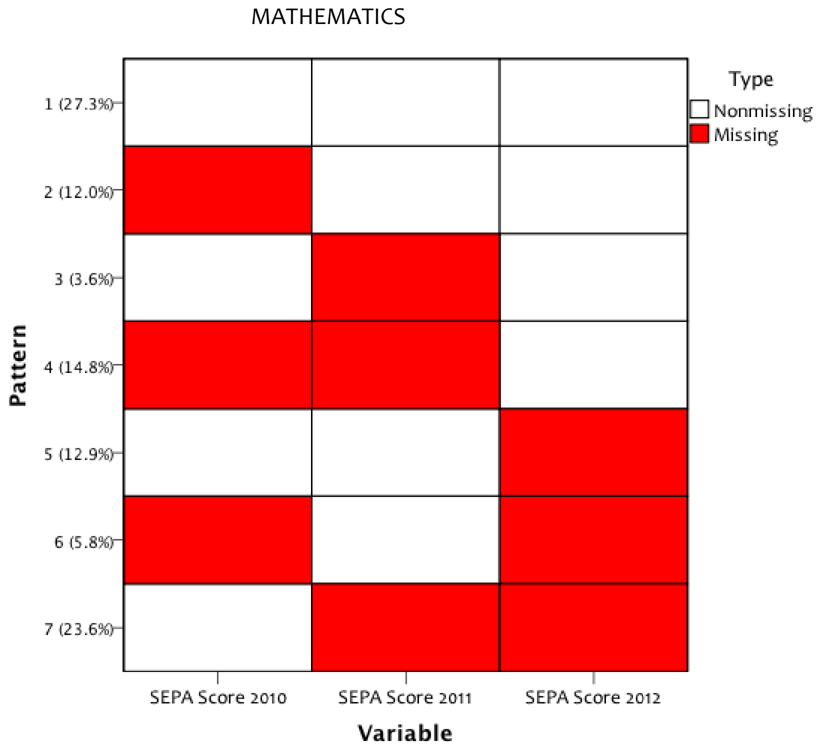


TABLE 30: TABULATED PATTERNS FOR STUDENT-LEVEL VARIABLES IN THE LANGUAGE DATABASE

Number of Cases	Missing Patterns[a]						Complete if ...[b]
	Gender	Age	Home Monthly Income	Mother's Educational Level	Number of Books at Home	Father's Educational Level	
15,400							15,400
224					X		15,624
669						X	16,069
3,665			X	X	X	X	20,373

Note: Patterns with less than 1% cases (204 or fewer) are not displayed. a. Variables are sorted on missing patterns. b. Number of complete cases if variables missing in that pattern (marked with X) are not used.

TABLE 31: TABULATED PATTERNS FOR STUDENT-LEVEL VARIABLES IN THE MATHEMATICS DATABASE

Number of Cases	Missing Patterns[a]						Complete if ...[b]
	Gender	Age	Mother's Educational Level	Home Monthly Income	Number of Books at Home	Father's Educational Level	
14,685							14,685
217					X		14,902
644						X	15,329
3,554			X	X	X	X	19,492

Note: Patterns with less than 1% cases (195 or fewer) are not displayed. a. Variables are sorted on missing patterns. b. Number of complete cases if variables missing in that pattern (marked with X) are not used.

TABLE 32: TABULATED PATTERNS FOR TEACHER-LEVEL VARIABLES IN THE LANGUAGE DATABASE

Number of cases	Missing Patterns[a]												Complete if …[b]
	Exper. 2011	**Gender 2011**	**Major 2011**	**ITT duration 2011**	**Exper. 2012**	**Gender 2012**	**Gender 2010**	**Exper. 2010**	**Major 2010**	**ITT duration 2010**	**Major 2012**	**ITT duration 2012**	
13,676													13,676
2,812											X	X	16,488
561			X	X					X	X	X	X	17,632
659					X	X					X	X	17,147
217							X	X	X	X	X	X	18,339
1,520							X	X	X	X			15,231

Note: Patterns with less than 1% cases (204 or fewer) are not displayed. a. Variables are sorted on missing patterns. b. Number of complete cases if variables missing in that pattern (marked with X) are not used.

TABLE 33: TABULATED PATTERNS FOR TEACHER-LEVEL VARIABLES IN THE MATHEMATICS DATABASE

Number of Cases	Missing Patterns[a]												Complete if …b
	Exper. 2011	**Gender 2011**	**Exper. 2012**	**Gender 2012**	**Major 2011**	**ITT duration 2011**	**ITT duration 2010**	**Major 2010**	**Exper. 2010**	**Gender 2010**	**Major 2012**	**ITT duration 2012**	
14,049													14,049
1,500							X	X	X	X			15,570
304							X	X	X	X	X	X	17,688
621					X	X	X	X			X	X	16,861
1,814											X	X	15,863
459			X	X							X	X	16,322

Note: Patterns with less than 1% cases (195 or fewer) are not displayed. a. Variables are sorted on missing patterns. b. Number of complete cases if variables missing in that pattern (marked with X) are not used.

Further analyses of patterns charts, that order variables and patterns to reveal monotonicity (i.e., the event that a variable is missing for a particular individual implies that all subsequent variables are missing for that individual), show that these datasets are nonmonotone. Indeed, the missing-data patterns observed can be classified as arbitrary, this is; missing values are haphazardly dispersed throughout the data. This has implication for potential missing data treatment, as the monotone imputation method[45] cannot be used for arbitrary missing data.

Missingness may be systematic despite the seemingly random pattern. The following section explores this issue by analysing the mechanisms of missing data.

Missing Data Mechanism

The missing data mechanism describes how an individual's propensity for missing data is related to other variables, if at all. Rubin's (1976) seminal work on missing data theory distinguishes between three mechanisms, namely:

- Missing Completely At Random (MCAR): The probability of missing data on Y is unrelated to other measured variables and is unrelated to the would-be values of Y itself. The observed scores are a random sample of the hypothetically complete data set.
- Missing At Random (MAR): The probability of missing data on Y is related to some other measured variable but is unrelated to the would-be values of Y itself. After controlling for other variables there is no association between the propensity of missing data on Y and the would-be values of Y.
- Not Missing at Random (NMAR): The probability of missing data on Y is related to the would-be values of Y itself (either directly or spuriously). NMAR is problematic and introduces bias when the would-be outcome scores determine missingness.

Thus, missing data mechanisms describe how the probability of a missing value on Y relates to other variables or to the would-be values of Y itself. Rubin's mechanisms act as assumptions for missing data analyses. List-wise deletion assumes MCAR. Modern

[45] This is, methods using Fully Conditional Specification Algorithms for multiple imputation.

approaches, such as maximum likelihood and multiple imputation, assume MAR. NMAR mechanisms, in turn, require specialized analysis models (e.g., selection models, pattern mixture models) (Collins, Schafer and Kam, 2001).

The data cannot differentiate MAR and NMAR because they posit different associations between missingness and the would-be values of Y. MCAR is the only mechanism with testable propositions. In order to test if data are missing completely at random, a missing data indicator was created for each incomplete variable and comparisons of group mean differences conducted on the variables, using t-tests and Cohen's d effect sizes. These separate t-tests allow the identification of variables whose pattern of missing values may be influencing the quantitative (scale) variables. The t-test is computed using an indicator variable that specifies whether a variable is present or missing for an individual case. Judging by the effect size of the differences in the quantitative variables between those cases with missing and valid values in the indicator variables, it appears that, in both databases, students whose parents have higher educational levels and that come from higher income households are less likely to have complete teacher-level data. In addition, teachers with fewer years of experience are more likely to have missing values in several variables associated with them.

These patterns are likely to be caused by limitations in the coverage of the secondary sources used to obtain information at the teacher level and are corroborated by the analysis of school sector (i.e., public, private subsidised, and private non-subsidised) versus indicator variables. This analysis shows that teacher-level data are more likely to be missing in private non-subsidised schools, institutions that operate autonomously and are not always legally obliged to submit information to the Ministry of Education.

The patterns presented above are indications that the data may not be missing completely at random. This conclusion can be confirmed through Little's MCAR test. The null hypothesis for Little's MCAR test is that the data are missing completely at random (MCAR). Four separate Little's MCAR tests were performed for each dataset (i.e., for the dependant variables, for the student-level variables, for the teacher-level variables and for the school-level variables). Because the significance value for all of the tests is less than 0.01[46], we can conclude that the data are not missing completely at random.

These diagnostic analyses show that the data is at least missing at random. As the theorem of ignorability can never be contradicted by the observed data, namely it is impossible to distinguish empirically between MAR and NMAR response mechanisms (Little and Rubin, 2002), it is assumed that the data is MAR.

[46] **Little's MCAR tests results for the language data set**
Dependent variables: Chi-Square = 446.606, DF = 9, Sig. = .000
Student-level variables: Chi-Square = 950.539, DF = 41, Sig. = .000
Teacher-level variables: Chi-Square = 4,713.256, DF = 52, Sig. = .000
School-level variables: Chi-Square = 697.282, DF = 11, Sig. = .000.
Little's MCAR tests results for the mathematics data set
Dependent variables: Chi-Square = 181.275, DF = 9, Sig. = .000
Student-level variables: Chi-Square = 918.446, DF = 41, Sig. = .000
Teacher-level variables: Chi-Square = 4,556.095, DF = 49, Sig. = .000
School-level variables Chi-Square = 694.117, DF = 11, Sig. = .000.

APPENDIX 3: PARAMETER CHAINS AND STANDARD MARKOV CHAIN MONTE CARLO CONVERGENCE DIAGNOSTICS FOR MODEL 8

FIGURE 56: PARAMETER CHAINS FOR MODEL 8 LANGUAGE

FIGURE 57: PARAMETER CHAINS FOR MODEL 8 MATHEMATICS

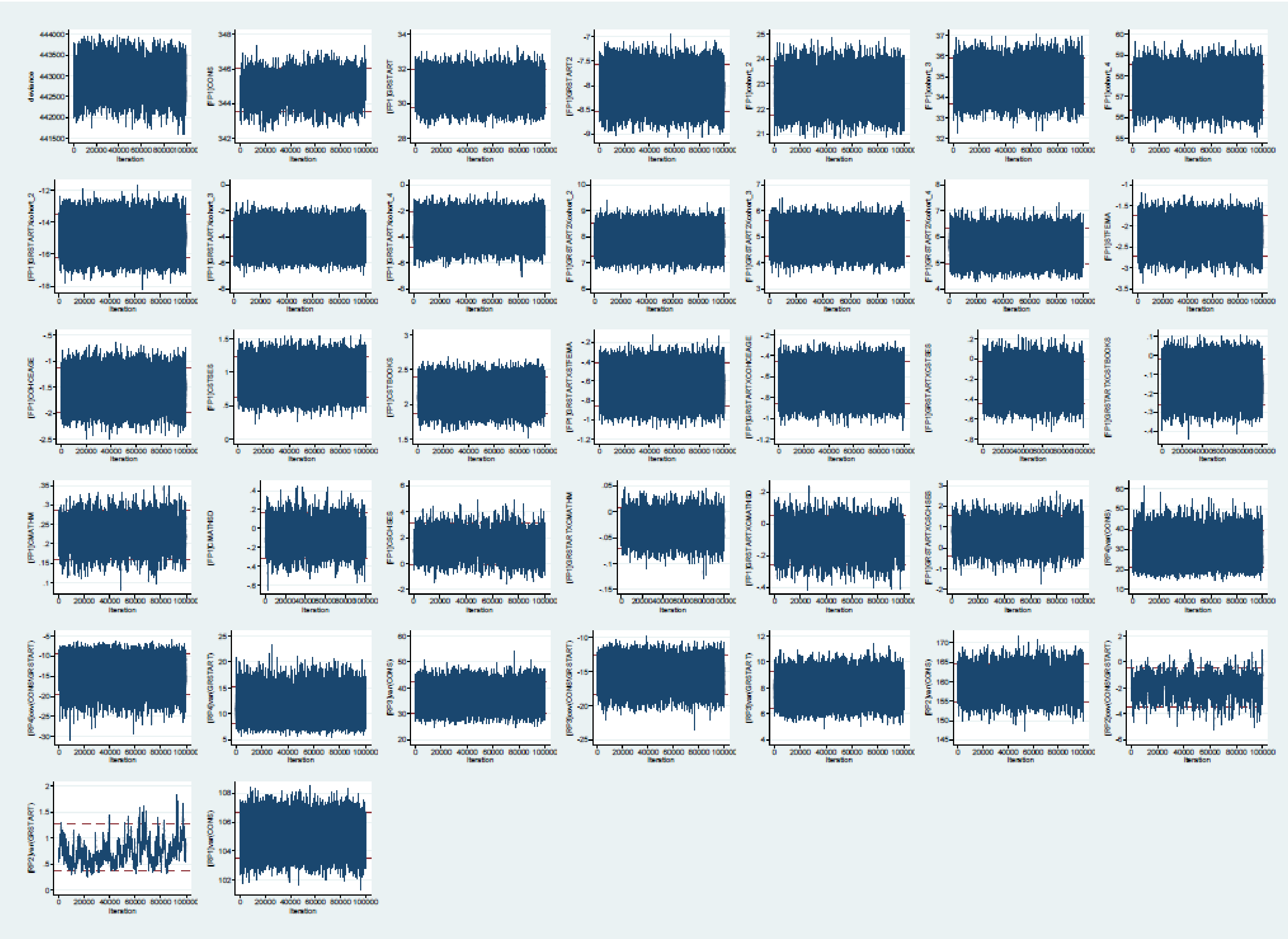
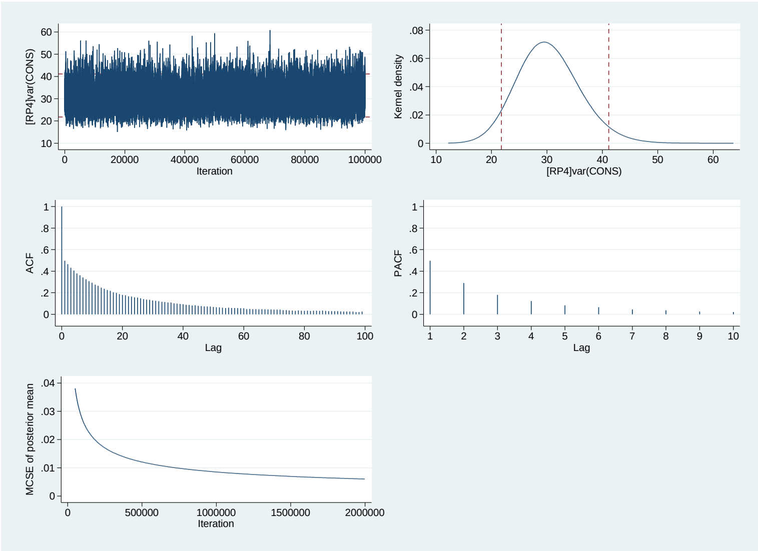


FIGURE 58: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR SCHOOL-LEVEL INTERCEPT VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

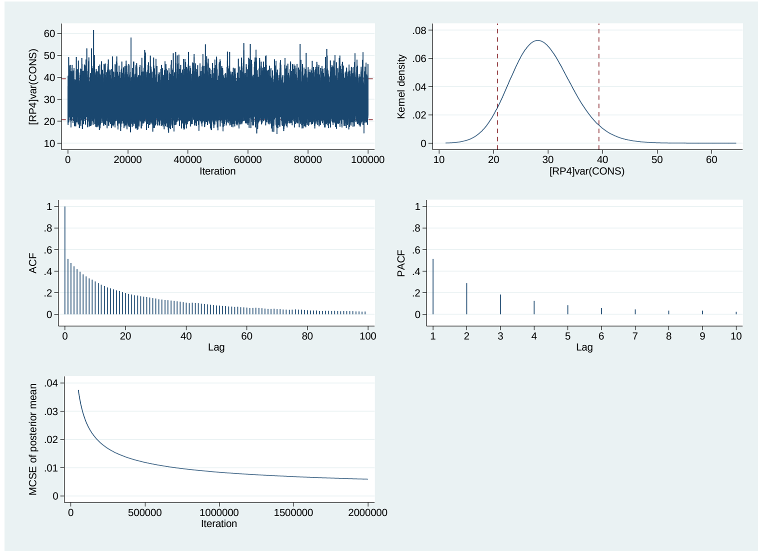
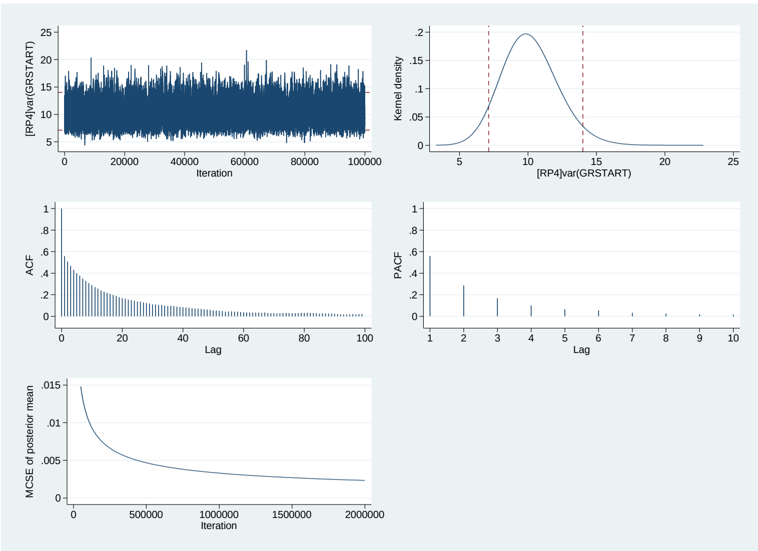


FIGURE 59: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR SCHOOL-LEVEL SLOPE VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

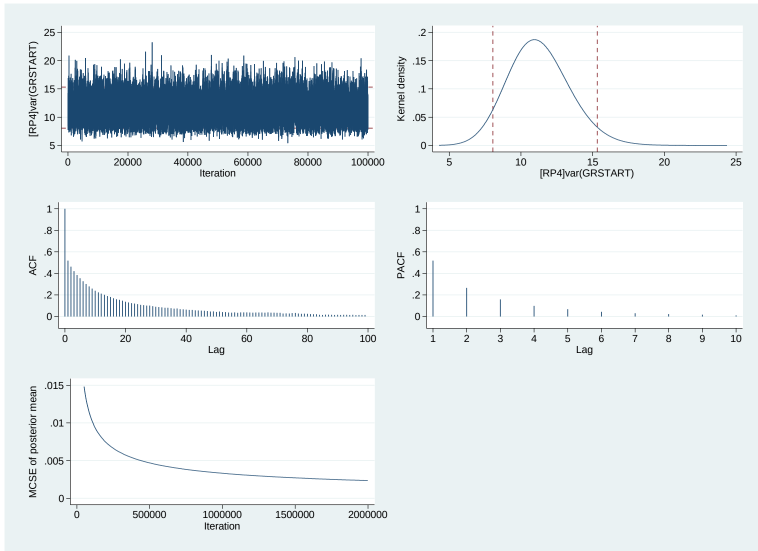
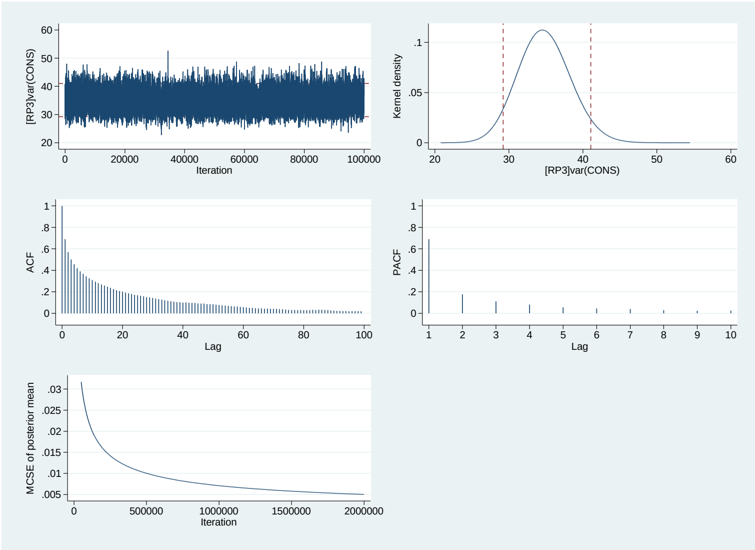


FIGURE 60: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR TEACHER-LEVEL INTERCEPT VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

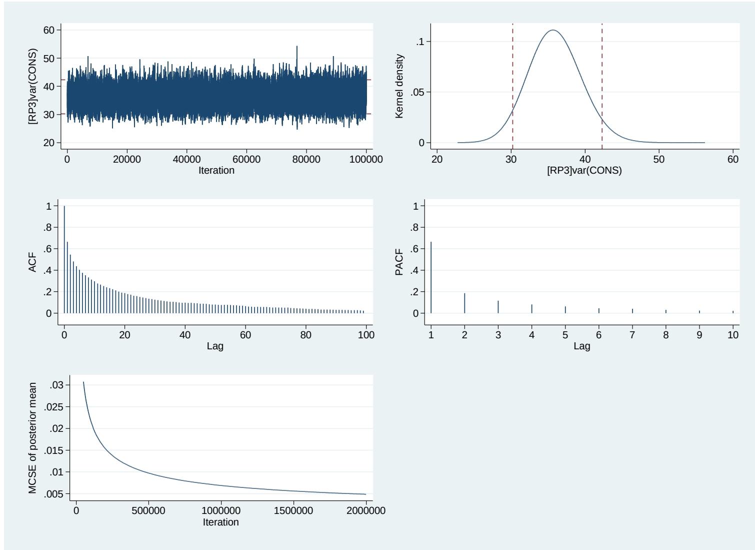
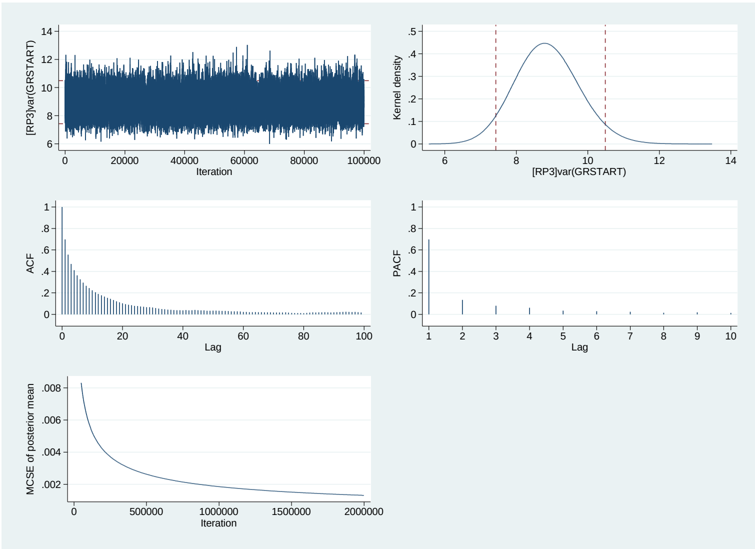


FIGURE 61: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR TEACHER-LEVEL SLOPE VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

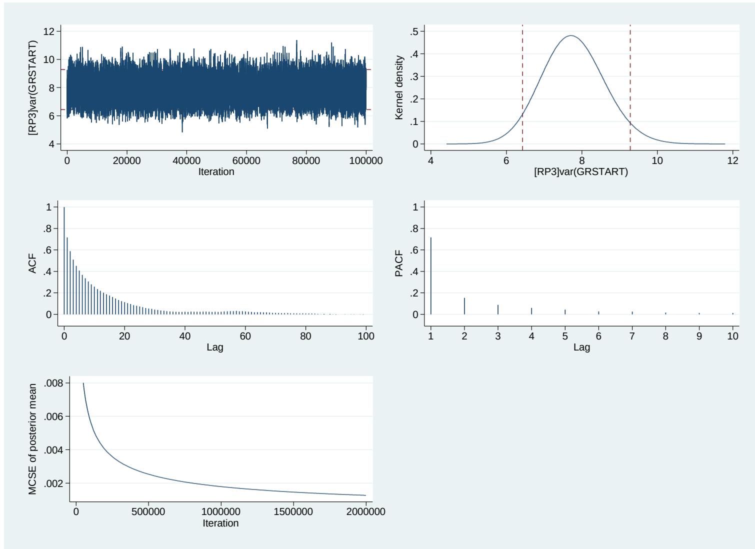
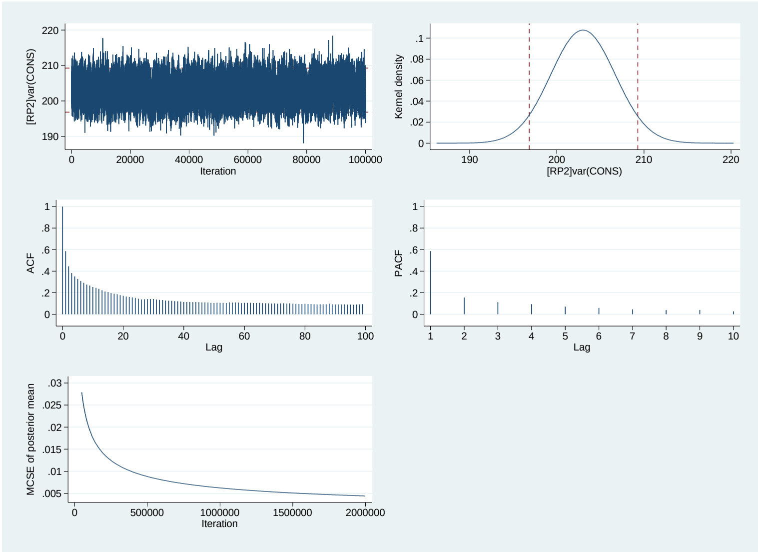


FIGURE 62: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR STUDENT-LEVEL INTERCEPT VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

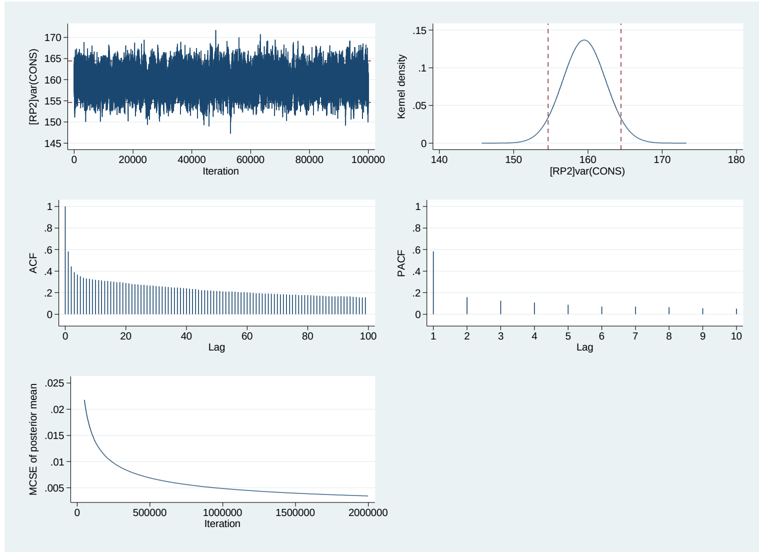
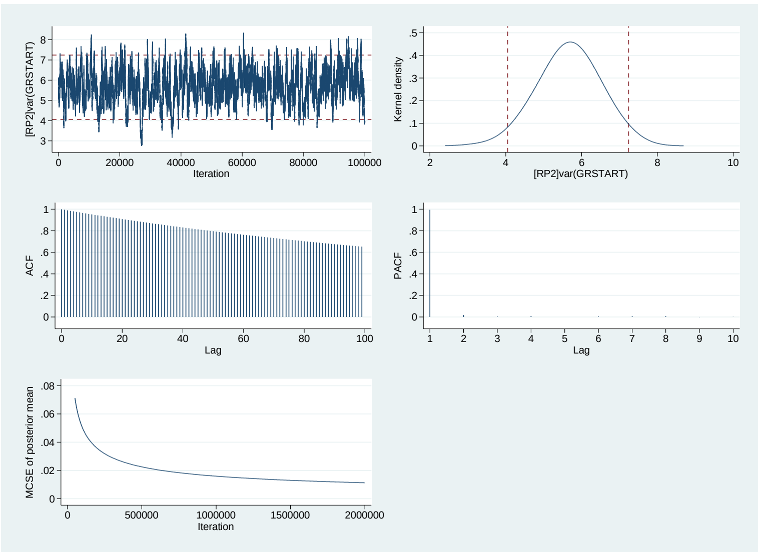


FIGURE 63: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR STUDENT-LEVEL SLOPE VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

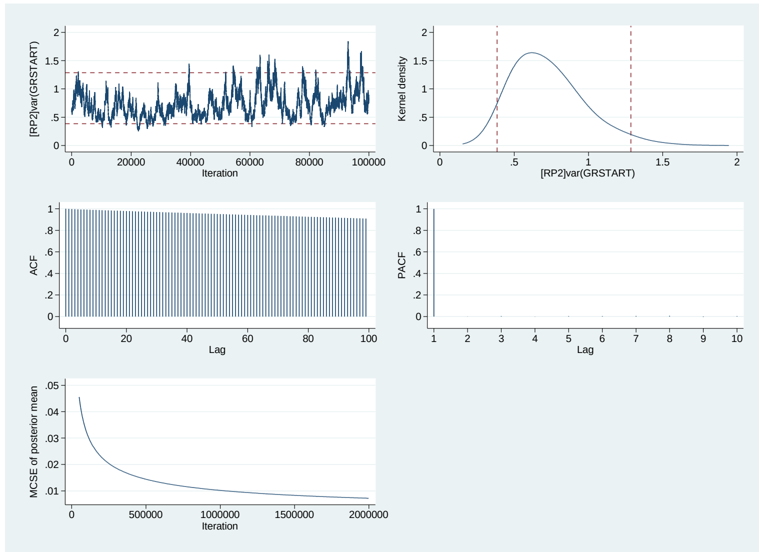
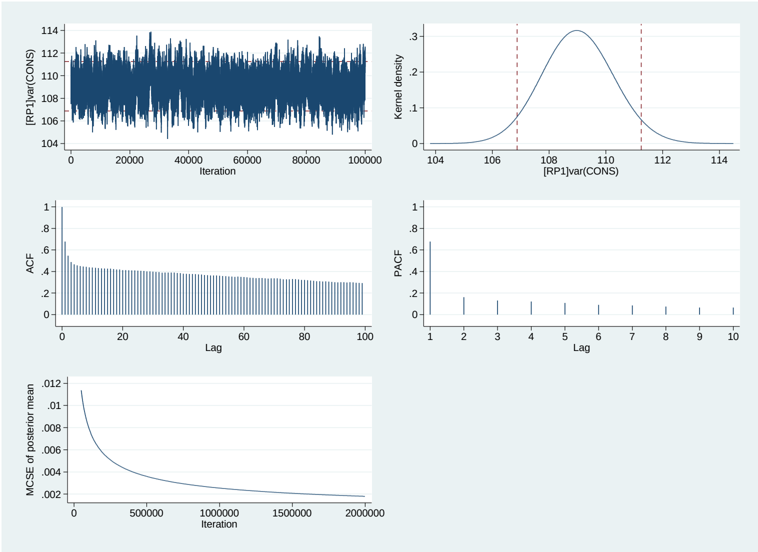


FIGURE 64: FIVE-WAY MCMC GRAPHICAL DIAGNOSTICS FOR MEASUREMENT-LEVEL INTERCEPT VARIANCE, MODEL 8

LANGUAGE



MATHEMATICS

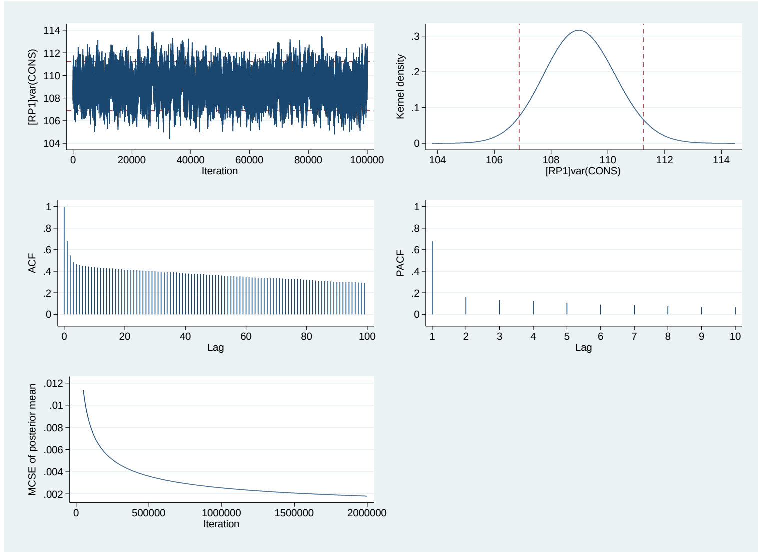


TABLE 34: MCMC CONVERGENCE DIAGNOSTICS FOR RANDOM PARAMETERS IN MODEL 8

	Raftery Lewis		Brooks Draper
	(2.5%)	(97.5%)	(mean)
LANGUAGE			
School Intercept Variance $\boldsymbol{\sigma^2_{u000}}$	10,562	9,654	1115
School Slope Variance $\boldsymbol{\sigma^2_{u100}}$	11,176	10,444	169
Teacher Intercept Variance $\boldsymbol{\sigma^2_{u00}}$	11,554	10,750	774
Teacher Slope Variance $\boldsymbol{\sigma^2_{u10}}$	10,924	10,954	5,312
Student Intercept Variance $\boldsymbol{\sigma^2_{u0}}$	10,050	9,792	6
Student Slope Variance $\boldsymbol{\sigma^2_{u1}}$	188,846	69,238	389,864
Occasion Intercept Variance $\boldsymbol{\sigma^2_{e}}$	10,752	11224	1
MATHEMATICS			
School Intercept Variance $\boldsymbol{\sigma^2_{u000}}$	10,580	9,390	1,082
School Slope Variance $\boldsymbol{\sigma^2_{u100}}$	10,372	9,724	170
Teacher Intercept Variance $\boldsymbol{\sigma^2_{u00}}$	11,744	10,942	728
Teacher Slope Variance $\boldsymbol{\sigma^2_{u10}}$	13,074	11,394	4,932
Student Intercept Variance $\boldsymbol{\sigma^2_{u0}}$	10,294	10,036	4
Student Slope Variance $\boldsymbol{\sigma^2_{u1}}$	187,306	199,322	159,213
Occasion Intercept Variance $\boldsymbol{\sigma^2_{e}}$	8,310	8,278	1

APPENDIX 4: OPERATIONAL DEFINITION OF CONSTRUCTS AND DESCRIPTIVE STATISTICS

The operational definition of the variables in the models fitted in Chapters 4 to 6 is presented below, followed by a table that summarizes the characteristics of the schools, teachers and students in the final sample by presenting descriptive statistics.

- **Outcome measures.** This study seeks to estimate educational effects on student achievement growth. Implementing such an approach requires longitudinal data, in which students are observed over multiple time points. The outcome measures for the present study are students' language and mathematics SEPA tests scores. The SEPA language and mathematics tests are designed to measure student performance on the grade-level competencies specified in the national curriculum standards and are, therefore, criterion-referenced assessments. The test scores in language and mathematics are both vertically and horizontally equated by the SEPA project analysts using item response theory (IRT), making it possible to compare scores among the different grades assessed and between the different time-points of assessment. Students' language and mathematics achievements were tested on up to five occasions; annually from 2007 to 2012. Of these assessment cycles, however only scores from 2010 to 2012 were equated. Thus, the SEPA project provides the data needed to measure the effect of schools and teachers on student achievement. Longitudinal data is key to enhance the validity of causal inferences by providing a basis for assessing the direction of causation between our predictors and outcome variables and by making possible to control, to some extent, selection effects (Cook and Campbell, 1979). In this study, student achievement in language and mathematics subjects are modelled separately giving way to two different sets of models.

All of the correlates used in this study are either input or context variables and were selected according to three criteria: to examine relationships often tested in the literature, to maximize explanatory power, and to avoid problems of multicollinearity. Also, given that multilevel models can rapidly become complex, attention was given to the principle of parsimony.

The correlates considered at the student-level are the following[47]:

- **Gender (*Female*).** The School Enrolment Recording System (SERS) provided complete information on student gender. The baseline of this variable is male.
- **Age (*Age*).** Students' dates of birth were retrieved from the School Enrolment Recording System (SERS) and transformed to reflect student age, calculated in years and months, as in December of 2010. The variable was cohort-mean centered.
- **Student socioeconomic status (*SES*).** This indicator was created by collapsing the variables 'average monthly household income', 'mother's educational level' and 'father's educational level'. The contribution of each variable was weighted based on results of exploratory factor analysis. The information was retrieved from the SIMCE parent questionnaire databases. For 'average monthly household income' there are thirteen possible categories ordered from the lowest to the highest range of income. Parents were also asked to indicate their highest completed educational level. Educational level goes from 1, meaning no formal education attended, to 20, referring to completed doctoral studies.
- **Number of books at home (*Books*).** This variable was also retrieved from the SIMCE parent questionnaire databases and consists of five categories (1=None, 2=Less than 10, 3=Between 10 and 50, 4=Between 51 and 100, 5=More than 100).

The school-level variables included in the models are:

- **School overall achievement (*Achievement Mean*).** The SIMCE test scores for students in Year 4, collected by the SIMCE office from the Chilean Ministry of Education, were used as an indicator of school overall achievement. The average of school scores for the years 2009, 2010 and 2011 were used to create this new indicator. When modelling student progress in language, this variable refers to the school average score in the three years in the SIMCE Reading Test. Correspondingly, when modelling student progress in mathematics this variable indicates the school average score in the SIMCE Mathematics Test.

[47] The variables' abbreviated name, used in the equations, appears here in parenthesis.

- **School Achievement standard deviation (*Achievement SD*)**. This variable refers to referring to the within-school standard deviation in the SIMCE test scores for the relevant subject as a measure of diversity in the levels of achievement of the student body.
- **Type of School** (***Private Subsidised*** and ***Private Non-Subsidised*****).** In terms of administration, there are three main categories of schools in Chile, namely, municipal (public) schools, private subsidized schools and private non-subsidised schools[48]. This distinction led to the creation of two dummy variables in our models, for which 'Public schools' is the baseline category. The information was retrieved from the SIMCE Assessment System databases.
- **School Socioeconomic Status (*School SES*).** This variable is a composite indicator, created and calculated by the Chilean Ministry of Education in the context of the SIMCE Assessment System. The variables considered in this indicator are: Mother's level of education, Father's level of education, Average monthly household income and the Social risk index of the school. Five ordered levels are defined from 1 to 5, level 1 being the lowest socio-economic level.
- **School Size *(School Size*)**. This variable was retrieved from the MINEDUC enrolment databases, which record the number of students in each school at the beginning of the academic year. The indicator was calculated as the average number of students at the primary level in the school for the years 2008, 2009 and 2010.
- **School Geographical Location (*Rural*).** This dichotomous variable distinguishes whether the school is located in a rural or urban zone. "Urban" is the baseline category in the models.

[48] Municipal (public) schools can be either Municipal DAEM schools or Municipal corporation schools. Municipal DAEM schools are those dependants of a department within the Municipality (called Departamentos de Administración de la Educación Municipal (DAEM). Municipal corporation schools are autonomous schools, constituted as a Corporation, where the Mayor is the head of the Board of the Corporation. The Corporation does not belong to the municipality and hence the statute of municipal public servants does not bind its personnel. Most Municipal schools are DAEM. Private subsidized and schools include religious schools, for profit lay schools and not for profit lay schools. Private non-subsidised schools finance themselves 100% from tuition payments, while Private subsidized finance themselves from the fiscal budget mostly and Municipal schools are free and financed themselves 100% from the fiscal budget.

Finally, all of the teacher-variables presented below, were accessed via the Teacher Census data sets.

- **Gender (*female teacher*).** This variable refers to teachers' gender and the baseline category is "male".
- **Years of teaching experience (*experience*).** The number of years that educators have been teaching in any school.
- **Initial teacher training programme duration (*itt duration*).** The length of initial teacher training programmes in Chile differs considerably, ranging from 1 to 6 years. This variable refers to the duration, in semesters, of the programmes in which the teachers in our sample were trained.
- **Major (*major*).** Teachers differ on whether they have undertaken specialized training in a particular subject or not. When modelling student progress in language, this variable refers to whether the training of the teacher associated to the student included a major in language. Correspondingly, if modelling student progress in mathematics, this variable indicates whether the teacher was trained in a programme including a major in mathematics. The baseline of these dichotomous variables refers to the condition without major.

Continuous variables were grand-mean centered[49] in order to ease interpretation, by providing a meaningful zero value, and avoid convergence problems (Enders and Tofighi, 2007; Kreft, De Leew and Aiken, 1995). In the following table, the descriptive statistics for school and teacher variables were calculated by weighting them by the number of students associated to them in the sample.

[49] This practice consists of subtracting the overall or grand mean from all values of a variable.

TABLE 35: OPERATIONAL DEFINITION AND DESCRIPTIVE STATISTICS OF VARIABLES IN CHAPTERS 4 TO 6[50]

CLASIFICATION	VARIABLE	DESCRIPTION	DEFINITION	MEAN	SD	MIN	MAX	SOURCE
Dependant Variables	SEPA Language Score 2010	SEPA 2010 standardised language test score	Continuous	376.697	29.124	275.0	476.2	SEPA
	SEPA Language Score 2011	SEPA 2011 standardised language test score	Continuous	398.301	27.500	297.2	493.8	SEPA
	SEPA Language Score 2012	SEPA 2012 standardised language test score	Continuous	410.250	27.401	316.2	504.8	SEPA
	SEPA Mathematics Score 2010	SEPA 2010 standardised mathematics test score	Continuous	370.650	28.647	272.0	469.8	SEPA
	SEPA Mathematics Score 2011	SEPA 2011 standardised mathematics test score	Continuous	391.698	30.610	229.4	498.2	SEPA
	SEPA Mathematics Score 2012	SEPA 2012 standardised mathematics test score	Continuous	406.824	33.672	297.0	517.2	SEPA
	N (Observations)			59,112				
Student-level variables	Gender	Student gender	Categorical 0 = Male, 1 = Female	0.499	-	-	-	SERS - STUDENT
	Age	Age, calculated in years and months, as in December of 2010	Continuous	10.664	1.278	8.4	16.7	SERS - STUDENT
	Average Monthly Household Income	13 possible categories ordered from the lowest to the highest range of income	Ordinal 1 = Lowest Income, to 13 = Highest range of income	5.456	4.244	1	13	SIMCE - STUDENT
	Mother's Educational Level	Mother's highest completed educational level.	Ordinal 1 = No formal education, to 20 = Completed doctoral studies	13.243	3.731	1	20	SIMCE - STUDENT
	Father's Educational Level	Father's highest completed educational level.	Ordinal 1 = No formal education, to 20 = Completed doctoral studies	13.412	3.868	1	20	SIMCE - STUDENT
	Student Socioeconomic Status	Composite variable based on monthly income', father's and mother's educational levels	Continuous	0	1	-2.807	1.969	SIMCE - STUDENT
	Number of Books at Home	Number of books at home	Ordinal 1=None, 2=Less than 10, 3=Between 10 and 50, 4=Between 51 and 100, 5=More than 100	3.093	1.077	1	5	SIMCE - STUDENT
	N (Students)			19,704				

[50] Continuous variables in the models were grand-mean centered before modelling. SD stands for standard deviation.

CLASIFICATION	VARIABLE	DESCRIPTION	DEFINITION	MEAN	SD	MIN	MAX	SOURCE
School-level variables	School Overall Reading Performance	Mean School SIMCE Reading Score	Continuous	271.031	26.965	207.00	319.33	SIMCE-SCHOOL
	School Reading Performance SD	School SIMCE Reading Score Standard Deviation	Continuous	44.010	4.922	28.41	54.06	SIMCE-SCHOOL
	School Overall Mathematics Performance	Mean School SIMCE Mathematics Score	Continuous	263.418	34.081	199.33	324.33	SIMCE-SCHOOL
	School Mathematics Performance SD	School SIMCE Mathematics Score Standard Deviation	Continuous	44.141	4.529	30.91	56.37	SIMCE-SCHOOL
	Private Subsidised	Dummy for private subsidised schools	Categorical 0 = Public, 1 = Private subsidised	0.401	-	-	-	SIMCE-SCHOOL
	Private Non-Subsidised	Dummy for private non-subsidised schools	Categorical 0 = Public, 1 = Private non-subsidised	0.222	-	-	-	SIMCE-SCHOOL
	School Socioeconomic Status	Composite indicator. The variables considered in this indicator are: Mother's level of education, Father's level of education, Average monthly household income and the Social risk index of the school.	Ordinal 1 = Lowest socio-economic level, to 5 = Highest socio-economic level	3.120	1.284	1	5	SIMCE-SCHOOL
	School Size	Average number of students at the primary level in the school for the years 2008, 2009 and 2010.	Continuous	881.340	409.675	59	2,247	SERS - SCHOOL
	School Geographical Location	Dichotomous variable indicating whether school is located in a rural area or not	Categorical 0 = Urban, 1 = Rural	0.092	-	-	-	SIMCE-SCHOOL
	N (Schools)			156				
Teacher-level variables								
Language								
	Teacher Gender		Categorical 0 = Male, 1 = Female	0.835	-	-	-	Teacher Census
	Years of Teaching Experience	Number of years educators have been teaching in any school	Continuous	17.845	12.001	0	44	Teacher Census
	Initial Teacher Training Programme Duration	Duration, in semesters, of the programmes in which the teacher was trained	Continuous	8.611	1.407	3	12	Teacher Census

CLASIFICATION	VARIABLE	DESCRIPTION	DEFINITION	MEAN	SD	MIN	MAX	SOURCE
	Major in the Subject	Whether the training of the teacher associated to the student included a major in language.	Categorical 0 = Without major, 1 = With major	0.325	-	-	-	Teacher Census
	N (Language Teachers)			851				
Mathematics								
	Teacher Gender		Categorical 0 = Male, 1 = Female	0.706	-	-	-	Teacher Census
	Years of Teaching Experience	Number of years educators have been teaching in any school	Continuous	18.865	12.487	0	47	Teacher Census
	Initial Teacher Training Programme Duration	Duration, in semesters, of the programmes in which the teacher was trained	Continuous	8.553	1.592	1	14	Teacher Census
	Major in the Subject	Whether the training of the teacher associated to the student included a major in language.	Categorical 0 = Without major, 1 = With major	0.382	-	-	-	Teacher Census
	N (Mathematics Teachers)			**812**				

APPENDIX 5: ANALYSIS OF CEILING EFFECTS

TABLE 36: PERCENTAGE OF STUDENTS AT EACH COHORT AND GRADE LEVEL THAT ACHIEVE THE MAXIMUM TEST SCORE, OVERALL AND BY SCHOOL SECTOR

	Grade Level					
	3	4	5	6	7	8
Cohort I						
LANGUAGE						
% Total	0.043	0.034	0.169	-	-	-
% Public	0.000	0.000	0.076	-	-	-
% Private Subsidised	0.052	0.000	0.000	-	-	-
% Private Non-Subsidised	0.101	0.134	0.479	-	-	-
MATHEMATICS						
% Total	0.093	0.069	0.337	-	-	-
% Public	0.000	0.000	0.000	-	-	-
% Private Subsidised	0.200	0.000	0.121	-	-	-
% Private Non-Subsidised	0.102	0.270	1.075	-	-	-
Cohort II						
LANGUAGE						
% Total	-	0.057	0.031	0.068	-	-
% Public	-	0.000	0.000	0.000	-	-
% Private Subsidised		0.000	0.000	0.000		
% Private Non-Subsidised	-	0.385	0.146	0.262	-	-
MATHEMATICS						
% Total	-	0.033	0.062	0.641	-	-
% Public	-	0.000	0.000	0.000	-	-
% Private Subsidised	-	0.000	0.000	0.000	-	-
% Private Non-Subsidised	-	0.198	0.295	2.484	-	-
Cohort III						
LANGUAGE						
% Total	-	-	0.038	0.080	0.035	-
% Public	-	-	0.000	0.000	0.000	-
% Private Subsidised	-	-	0.211	0.128	0.000	-
% Private Non-Subsidised	-	-	0.000	0.139	0.119	-
MATHEMATICS						
% Total	-	-	0.039	0.321	0.035	-

	Grade Level					
	3	4	5	6	7	8
% Public	-	-	0.000	0.000	0.000	-
% Private Subsidised	-	-	0.229	0.000	0.000	-
% Private Non-Subsidised	-	-	0.000	1.132	0.119	-
Cohort IV						
LANGUAGE						
% Total	-	-	-	0.031	0.037	0.041
% Public	-	-	-	0.000	0.000	0.000
% Private Subsidised	-	-	-	0.000	0.000	0.000
% Private Non-Subsidised	-	-	-	0.131	0.149	0.169
MATHEMATICS						
% Total	-	-	-	0.062	0.037	0.040
% Public	-	-	-	0.000	0.000	0.000
% Private Subsidised	-	-	-	0.000	0.000	0.000
% Private Non-Subsidised	-	-	-	0.260	0.150	0.169

Note: Dashes indicate that data are not available in that grade level for that cohort. Calculated using sample prior to performing multiple imputation.

TABLE 37: DISTRIBUTION INDICATORS FOR SEPA SCORES BY GRADE LEVEL AND COHORT, FINAL SAMPLE

	Grade Level					
	3	4	5	6	7	8
Cohort I						
N	5,782	5,782	5,782	-	-	-
Skewness Language (SE)	0.388 (0.036)	0.284 (0.034)	0.206 (0.042)	-	-	-
Kurtosis Language (SE)	-0.148 (0.076)	-0.126 (0.066)	-0.202 (0.082)	-	-	-
Skewness Mathematics (SE)	0.376 (0.041)	0.289 (0.039)	0.519 (0.054)	-	-	-
Kurtosis Mathematics (SE)	0.159 (0.070)	0.040 (0.071)	0.548 (0.134)	-	-	-
Cohort II						
N	-	5,108	5,108	5,108	-	-
Skewness Language (SE)	-	0.469 (0.038)	0.349 (0.049)	0.098 (0.045)	-	-
Kurtosis Language (SE)	-	-0.140 (0.07)	-0.139 (0.112)	-0.317 (0.105)	-	-
Skewness Mathematics (SE)	-	0.426 (0.035)	0.538 (0.051)	0.653 (0.039)	-	-
Kurtosis Mathematics (SE)	-	0.073 (0.079)	-0.063 (0.077)	0.455 (0.079)	-	-
Cohort III						
N	-	-	4,269	4,269	4,269	-
Skewness Language (SE)	-	-	0.405 (0.052)	0.233 (0.052)	0.237 (0.047)	-
Kurtosis Language (SE)	-	-	-0.342 (0.090)	-0.330 (0.083)	-0.444 (0.097)	-
Skewness Mathematics (SE)	-	-	0.565 (0.053)	0.529 (0.045)	0.525 (0.046)	-
Kurtosis Mathematics (SE)	-	-	-0.072 (0.130)	0.147 (0.083)	0.218 (0.103)	-
Cohort IV						
N	-	-	-	4,545	4,545	4,545
Skewness Language (SE)	-	-	-	0.244 (0.046)	0.395 (0.039)	0.205 (0.043)
Kurtosis Language (SE)	-	-	-	-0.452 (0.088)	-0.331 (0.076)	-0.300 (0.104)
Skewness Mathematics (SE)	-	-	-	0.603 (0.038)	0.542 (0.054)	0.429 (0.037)
Kurtosis Mathematics (SE)	-	-	-	-0.021 (0.074)	0.088 (0.098)	-0.064 (0.079)

Note: Dashes indicate that data are not available in that grade level for that cohort. Statistics obtained by combining the results across the 5 imputed data sets using Rubin's rules.

APPENDIX 6: MODEL 2 RESULTS, BY SCHOOL SECTOR

TABLE 38: FULL MODEL 2 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR LANGUAGE, BY SCHOOL SECTOR

	MODEL 2 LANGUAGE								
	PUBLIC SCHOOLS			PRIVATE SUBSIDISED SCHOOLS			PRIVATE NON-SUBSIDISED SCHOOLS		
	M (SE)	*p* value	95% CI	M (SE)	*p* value	95% CI	M (SE)	*p* value	95% CI
Fixed Part									
Intercept	338.594 (0.84)	<0.001	336.938 - 340.250	350.796 (1.368)	<0.001	348.090 - 353.502	365.556 (1.274)	<0.001	362.996 - 368.116
Time	34.628 (0.872)	<0.001	32.850 - 36.406	34.448 (1.881)	<0.001	29.709 - 39.188	34.391 (1.940)	<0.001	29.816 - 38.966
$Time^2$	-6.873 (0.381)	<0.001	-7.662 - -6.083	-8.168 (1.006)	0.001	-10.821 - -5.515	-8.078 (0.838)	<0.001	-10.085 - -6.071
Cohort 2	26.309 (0.58)	<0.001	25.169 - 27.448	24.204 (0.728)	<0.001	22.759 - 25.649	24.802 (0.966)	<0.001	22.898 - 26.706
Cohort 3	37.188 (0.629)	<0.001	35.950 - 38.427	36.098 (1.009)	<0.001	33.940 - 38.257	36.676 (0.892)	<0.001	34.925 - 38.427
Cohort 4	53.015 (0.594)	<0.001	51.850 - 54.180	56.269 (0.919)	<0.001	54.372 - 58.165	54.575 (0.944)	<0.001	52.718 - 56.432
Time x Cohort 2	-11.839 (0.976)	<0.001	-13.789 - -9.889	-8.006 (1.504)	0.001	-11.433 - -4.579	-9.194 (1.387)	<0.001	-11.955 - -6.432
Time x Cohort 3	-9.241 (1.120)	<0.001	-11.527 - -6.955	-8.376 (2.380)	0.014	-14.301 - -2.452	-12.807 (1.755)	<0.001	-16.542 - -9.072
Time x Cohort 4	-13.942 (0.996)	<0.001	-15.925 - -11.960	-15.471 (1.725)	<0.001	-19.504 - -11.439	-10.360 (1.974)	<0.001	-14.725 - -5.996
$Time^2$ x Cohort 2	3.648 (0.460)	<0.001	2.732 - 4.564	2.691 (0.762)	0.008	0.920 - 4.461	2.108 (0.639)	0.001	0.844 - 3.372
$Time^2$ x Cohort 3	2.193 (0.559)	0.001	1.034 - 3.352	2.847 (1.083)	0.040	0.174 - 5.520	3.644 (0.843)	0.001	1.848 - 5.441
$Time^2$ x Cohort 4	6.489 (0.458)	<0.001	5.584 - 7.394	7.184 (0.849)	<0.001	5.175 - 9.193	4.635 (1.006)	0.001	2.364 - 6.907
Random Part									
Random Level 2 (School)									
Intercept	25.218 (6.778)	-	11.503 - 38.933	65.518 (16.309)	-	32.409 - 98.628	26.569 (9.035)	-	8.742 - 44.396
Time/Intercept	-7.091 (3.166)	-	-13.702 - -0.481	-13.464 (7.502)	-	-30.399 - 3.470	-5.283 (4.110)	-	-13.701 - 3.135
Time	7.643 (2.621)	-	1.815 - 13.471	10.970 (2.858)	-	5.133 - 16.807	7.373 (2.702)	-	1.920 - 12.826
Random Level 2 (Student)									
Intercept	170.980 (5.485)	-	159.958 - 182.003	215.538 (5.395)	-	204.937 - 226.138	261.678 (8.700)	-	244.542 - 278.813
Time/Intercept	-3.023 (2.387)	-	-7.939 - 1.892	-9.414 (2.436)	-	-14.474 - -4.353	-17.919 (3.245)	-	-24.328 - -11.511
Time	2.899 (1.429)	-	0.049 - 5.748	1.635 (1.266)	-	-0.854 - 4.125	3.891 (1.963)	-	0.034 - 7.747
Random Level 1 (Occasion)									
Intercept	108.844 (2.307)	-	104.071 - 113.616	123.471 (3.192)	-	116.242 - 130.701	127.625 (3.189)	-	121.194 - 134.057
Student Initial Status – Growth Correlation	-0.136			-0.502			-0.562		
School ICC Initial Status	0.129			0.233			0.092		
School ICC Growth	0.725			0.870			0.655		
Units: Schools	64			60			32		
Units: Students	7,437			7,895			4,372		
Units: Occasions	22,311			23,685			13,116		
DIC	174,289			188,490			104,692		
pD	6,657			6,698			3,947		

TABLE 39: FULL MODEL 2 - POSTERIOR MEANS, STANDARD DEVIATIONS AND 95% CREDIBLE INTERVALS FOR MATHEMATICS, BY SCHOOL SECTOR

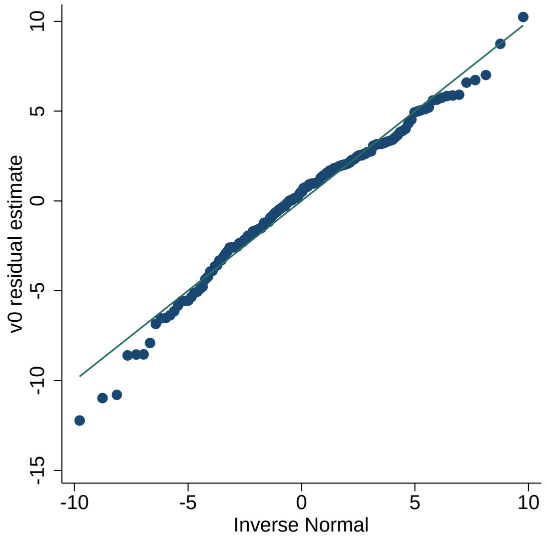
	MODEL 2 MATHEMATICS								
	PUBLIC SCHOOLS			PRIVATE SUBSIDISED SCHOOLS			PRIVATE NON-SUBSIDISED SCHOOLS		
	M (SE)	*p* value	95% CI	M (SE)	*p* value	95% CI	M (SE)	*p* value	95% CI
Fixed Part									
Intercept	335.076 (0.862)	<0.001	333.374 - 336.778	345.463 (1.583)	<0.001	342.340 - 348.587	357.611 (1.377)	<0.001	354.874 - 360.348
Time	30.623 (0.871)	<0.001	28.832 - 32.414	28.973 (3.891)	0.001	18.425 - 39.521	33.817 (2.264)	<0.001	28.283 - 39.350
Time^2	-8.097 (0.428)	<0.001	-9.029 - -7.165	-7.444 (2.107)	0.023	-13.227 - -1.660	-8.844 (1.070)	<0.001	-11.543 - -6.145
Cohort 2	23.371 (0.525)	<0.001	22.339 - 24.404	21.831 (0.774)	<0.001	20.244 - 23.418	24.180 (0.976)	<0.001	22.242 - 26.118
Cohort 3	29.452 (0.580)	<0.001	28.304 - 30.599	35.031 (0.761)	<0.001	33.506 - 36.556	37.700 (0.955)	<0.001	35.793 - 39.606
Cohort 4	51.816 (0.662)	<0.001	50.453 - 53.179	56.266 (0.879)	<0.001	54.442 - 58.089	62.512 (0.897)	<0.001	60.752 - 64.273
Time x Cohort 2	-21.896 (1.062)	<0.001	-24.092 - -19.701	-12.501 (2.978)	0.010	-20.311 - -4.691	-14.373 (2.679)	0.001	-20.858 - -7.889
Time x Cohort 3	-5.682 (1.286)	0.001	-8.463 - -2.900	-6.932 (3.303)	0.094	-15.596 - 1.732	-5.606 (2.376)	0.049	-11.174 - -0.039
Time x Cohort 4	-3.636 (1.371)	0.024	-6.682 - -0.591	-5.021 (3.739)	0.243	-14.988 - 4.947	-5.682 (2.383)	0.047	-11.272 - -0.092
Time^2 x Cohort 2	11.089 (0.530)	<0.001	9.979 - 12.198	6.871 (1.472)	0.007	2.994 - 10.747	7.343 (1.296)	0.001	4.196 - 10.490
Time^2 x Cohort 3	6.618 (0.599)	<0.001	5.331 - 7.905	6.337 (1.802)	0.020	1.547 - 11.126	4.666 (1.202)	0.006	1.809 - 7.524
Time^2 x Cohort 4	6.662 (0.664)	<0.001	5.183 - 8.141	6.554 (1.921)	0.023	1.400 - 11.709	5.895 (1.294)	0.004	2.758 - 9.032
Random Part									
Random Level 2 (School)									
Intercept	27.881 (6.914)	-	14.124 - 41.639	92.553 (20.383)	-	52.152 - 132.955	39.249 (15.849)	-	6.261 - 72.238
Time/Intercept	-8.474 (2.896)	-	-14.255 - -2.693	-19.931 (8.186)	-	-37.700 - -2.163	-8.305 (6.778)	-	-23.279 - 6.668
Time	7.780 (2.230)	-	3.090 - 12.470	11.144 (3.638)	-	3.155 - 19.133	8.480 (3.288)	-	1.701 - 15.259
Random Level 2 (Student)									
Intercept	123.026 (3.456)	-	116.237 - 129.815	178.528 (5.169)	-	168.078 - 188.978	237.287 (7.650)	-	222.252 - 252.321
Time/Intercept	0.363 (1.163)	-	-1.961 - 2.687	-6.349 (1.894)	-	-10.384 - -2.315	-0.290 (2.189)	-	-4.586 - 4.007
Time	0.842 (0.249)	-	0.353 - 1.330	0.627 (0.214)	-	0.205 - 1.048	0.911 (0.942)	-	-0.937 - 2.759
Random Level 1 (Occasion)									
Intercept	100.245 (1.497)	-	97.182 - 103.308	117.163 (2.227)	-	112.178 - 122.148	125.505 (3.177)	-	118.643 - 132.367
Student Initial Status – Growth Correlation	0.036			-0.600			-0.020		
School ICC Initial Status	0.185			0.341			0.142		
School ICC Growth	0.902			0.947			0.903		
Units: Schools	64			60			32		
Units: Students	7,437			7,895			4,372		
Units: Occasions	22,311			23,685			13,116		
DIC	172,211			186,199			104,109		
pD	6,062			6,578			3,834		

Note: Blocks of rows in the tables show relevant parameter estimates and residual variances at the measurement occasions, student and school levels. The first column for each cohort shows the posterior means and standard deviations (in parenthesis) of the fixed and random-part parameters (analogous to frequentist parameter estimates and standard errors). In the second column, 95% CIs are presented, calculated as the 2.5th and 97.5th percentiles of the MCMC chain for each parameter.

APPENDIX 7: NORMAL PROBABILITY PLOTS FOR RESIDUALS

FIGURE 65: PLOT OF LEVEL 3 (SCHOOL) STANDARISED RESIDUALS IN MODEL 4 FOR LANGUAGE, AGAINST NORMAL SCORES

INTERCEPT



SLOPE

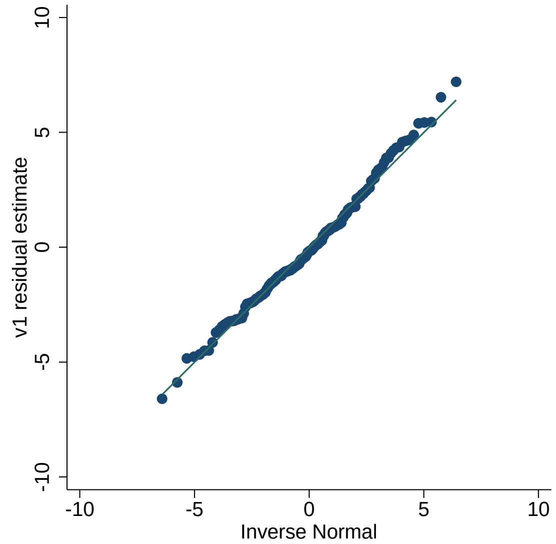
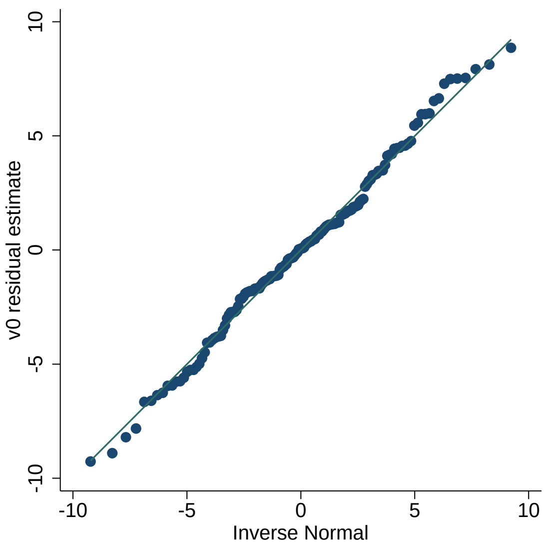


FIGURE 66: PLOT OF LEVEL 3 (SCHOOL) STANDARISED RESIDUALS IN MODEL 4 FOR MATHEMATICS, AGAINST NORMAL SCORES

INTERCEPT



SLOPE

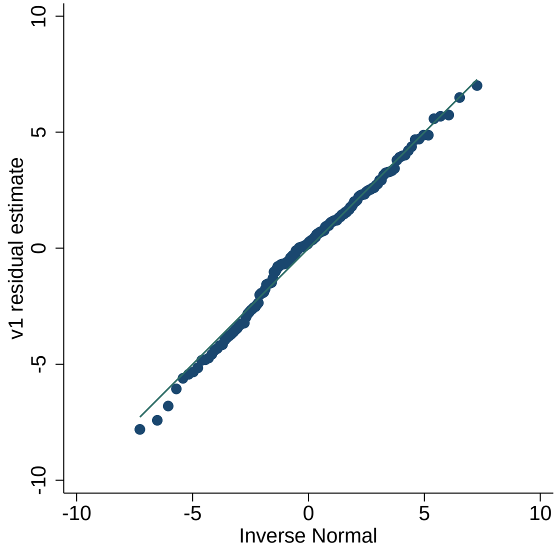


FIGURE 67: PLOT OF LEVEL 3 (TEACHER) STANDARISED RESIDUALS IN MODEL 6 FOR LANGUAGE, AGAINST NORMAL SCORES

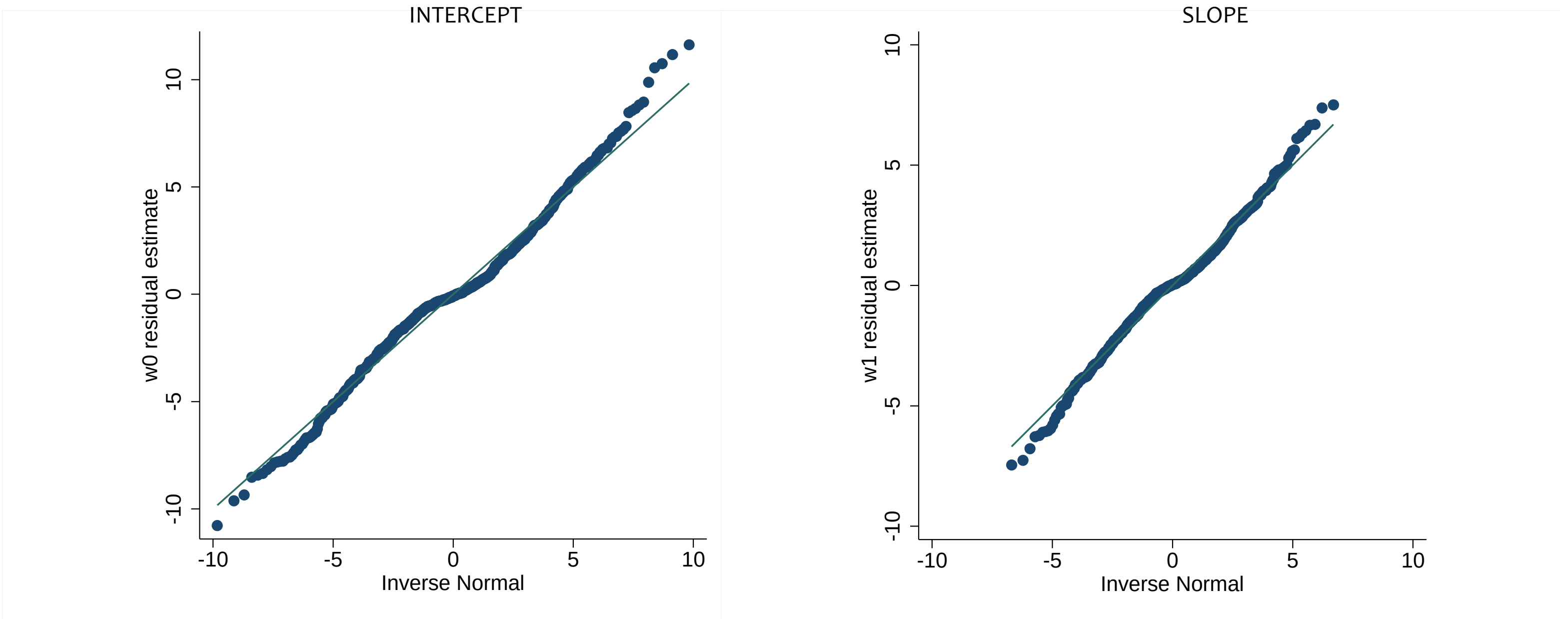


FIGURE 68: PLOT OF LEVEL 3 (TEACHER) STANDARISED RESIDUALS IN MODEL 6 FOR MATHEMATICS, AGAINST NORMAL SCORES