BEYOND THE TRADITIONAL SCHOOL VALUE-ADDED APPROACH: ANALYSING COMPLEX MULTILEVEL MODELS TO INFORM EXTERNAL AND INTERNAL SCHOOL ACCOUNTABILITY IN CHILE

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2015

PATRICIO EDUARDO TRONCOSO RUIZ

SCHOOL OF SOCIAL SCIENCES

Contents

Word count: 80,745

Tables

Figures

ABSTRACT

In the last few decades, educational research has largely demonstrated the effects of the socio-economic background on academic performance. Traditionally, researchers have used the so-called contextualised value-added (CVA) concept, implemented via multilevel statistical models, to assess variation in learning outcomes arising from schools and pupils. Depending on the stakeholders they intend to inform, two basic types of CVA models can be defined: models for school accountability and models for school choice. School accountability models can be further distinguished according to the 'recipient' of the information: internal models provide information for school authorities to improve their own practices, while external models provide information for government officials to assess school performance for policy-making purposes. Despite the evidence in favour of the use of more complex models for school accountability, government practice in Chile has been restricted to the use of raw school averages in standardised tests as indicators of effectiveness, which have been used indiscriminately for the purposes of school accountability and school choice.

Using data from the Chilean National Pupil Database (SIMCE 2004-2006), this thesis demonstrates how the traditional CVA (2-level) models fall short in addressing the complex phenomenon of academic performance, especially in the context of a developing and highly unequal country, such as Chile. The novelty of the CVA modelling in this thesis is that it extends and improves the traditional models insofar as they explicitly assess the variation between pupils, classrooms, primary schools, secondary schools and local authorities, as well as the correlation between Mathematics and Spanish Language at all levels. This is done by implementing two univariate 4-level CVA models for progress in Mathematics and Spanish fitted separately via maximum likelihood estimation (MLE) and a bivariate 5-level cross-classified CVA model for progress in both subjects fitted via Markov Chain Monte Carlo (MCMC) estimation.

External school accountability measures were derived from the extended univariate and multivariate models and compared to measures derived from a model akin to the traditional approach. A number of key differences were found, leading to the conclusion that further adjustments to the traditional CVA models are not negligible. The univariate 4-level CVA models provide more insight into school accountability than the traditional approach in a straightforward fashion, while the bivariate 5-level model encompasses a more reliable and ultimately comprehensive view on school performance.

With regard to internal school accountability, further models were specified with the purpose of analysing pupils' heterogeneity to inform school improvement processes. The concept of "cultural capital" (Bourdieu, 1977) was chosen to shed light on the matter. Since cultural capital is essentially immeasurable, a latent variable was constructed from a group of manifest variables related to access and use of reading materials. From a substantive point of view, this thesis shows how access to all sorts of reading materials and reading habits can have not only a relevant impact on pupils' progress in Language, but also in Mathematics.

Finally, this thesis concludes around three main ideas: firstly, school value-added models for school accountability, either external or internal, need to take into account the complexity of influences affecting pupils' academic progress as thoroughly as possible, in order to make a fair assessment of schools' performance and/or to inform school improvement policies. Secondly, school effectiveness is not a unidimensional process, which implies that school value-added models should ideally (when there are available data) reflect upon the multidimensionality of the phenomenon and take into consideration the relationship between different subjects, as well as non-academic outcomes. Thirdly, CVA models can also be used to inform internal school accountability by analysing the effects of meaningful modifiable factors and potentially serve as drivers of school improvement policies.

DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses.

ACKNOWLEDGMENTS

This thesis would not have been possible without the contribution of many people in various aspects of this research process. Some have contributed actively and knowingly in key aspects throughout my whole PhD studies; while others have made somewhat smaller, but significant, contributions, sometimes even unknowingly.

First and foremost, I would like to express my deepest gratitude to my supervisors Pr Wendy Olsen and Dr Maria Pampaka, who have given me all the necessary guidance, knowledge and encouragement to finish this PhD. Their feedback has been crucial for the development of my research and myself as a researcher.

CMist has also played an enormous role in my road to PhD completion. Back in 2011, when I joined the CCSR (now CMist), I could have not imagined how far I was going to progress in my professional development. All the knowledge and experience I acquired through several statistical and methods courses run by CMist staff, in which I took part either as TA or as student, is invaluable. This is almost entirely responsibility of the CMist staff, especially Pr Tarani Chandola, Pr Natalie Shlomo and Dr Johan Koskinen. I would also thank Philippa Walker, who always responded promptly to all my innumerable queries regarding administrative matters. I would also like to thank the useful insights and advice from my internal assessor, Dr Mark Tranmer, who helped through my annual and mid-year reviews, as well as in many other less formal occasions.

I would also like to mention my fellow PhD colleagues in CMist with whom I have shared many seminars, training courses, conferences, coffee breaks, etc. I would also like to acknowledge the contribution of my "conference buddy" Adrián Leguina (now Dr), with whom I had numerous conversations about various "statistically significant" topics. A special mention goes to my personal and professional companion in life Ana Morales (Anita), who very patiently listened to me go on and on about my research and gave me invaluable statistical, methodological and theoretical advice throughout my PhD: I could not thank you enough!

Family, of course, is an inexhaustible source of motivation, encouragement and more importantly, endearment. My wife Anita and my daughter Amanda are the ones who have witnessed all the good and not-so-good moments in this PhD, and have always been the first ones to provide the precise words and affection I needed to get by. Thank you both for being adventurous enough to embark on this long haul travel. The three of us have done our own unique contribution to this adventure with the conviction that no matter the outcome, either fail or succeed, we are all in this together. We have come a long way!

On the other side of the pond, I would like to thank my family of origin back in Chile, my mother Marisol, my brother Víctor and my sister Marisol, as well as my family in-law, my mother-in-law Leonor, my father-in-law Alejandro, my brother-in-law Alejandro (and his own family), my sister-in-law Verónica (and little Gustavín). This PhD is also yours. Thanks for giving me the courage, determination and motivation to pursue my dreams (and Anita's).

On this side of the pond, it would be unfair not to thank the Chilean community in Manchester, especially my good friends: Dr Christian Salas, Darinka Radovic, Andrés Olivares, Natalia Zúñiga, Hugo Benítez, Fiorella Vallebuona and others, who have, pretty much inadvertently, contributed in various ways. Also, and more importantly, they have given me the support and good times much needed to see the PhD through.

Finally, this thesis would have not been possible without the institutional support of the Chilean Government. Firstly, through the CONICYT-Becas Chile Scholarship Programme, which funded my PhD studies, and secondly, through the Chilean Ministry of Education (MINEDUC), which provided the data sets (SIMCE) used in this research. Despite the relevance of this institutional support, this research was conducted independently and the views expressed here are held by the author, not to be attributed to others.

Behind this thesis, not only are there countless hours of work, but also countless hours of stress, frustration, joy, satisfaction, etc. However, behind those countless sentiments, there are only two pillars that kept me standing throughout this journey.

In short, this thesis is dedicated to you, Anita and Amanda.

THE AUTHOR

Patricio Troncoso is currently a PhD student in Social Statistics at the University of Manchester. Since 2011, he has been researching on school effectiveness in Chilean schools. His research is focused on the methodological challenges of analysing school value-added on Mathematics and Language attainment using complex multilevel models as well as substantively focusing on the socio-economic inequalities of the Chilean education system. His research interests are mainly related to the broad areas of Statistics and Education with an emphasis on Multilevel Modelling and educational inequalities. Before starting his PhD, Patricio graduated from the University of La Frontera (Chile) with a Bachelor's degree in Sociology (2007) and graduated from the University of Palermo (Italy) and the University of Deusto (Spain) with a Master's degree in Public Policy (2010).

From the work presented in this thesis, the following journal article has been published:

Troncoso, P., Pampaka, M., & Olsen, W. (2015). Beyond traditional school value-added models: a multilevel analysis of complex school effects in Chile. School Effectiveness and School Improvement, 1–22. DOI: 10.1080/09243453.2015.1084010

Chapter 1: Introduction

In the last few decades, educational research has largely demonstrated the effects of socio-economic and demographic characteristics of the schools and the pupils on academic performance. Traditionally, educational researchers have implemented the so-called contextualised value-added (CVA) models to assess variation in learning outcomes (usually standardised tests of Mathematics, Language and other subjects to a lesser extent) arising from schools and pupils, by implementing multilevel statistical models. These models can also be used (and have been used) for informing school accountability to various stakeholders, including policy-makers, head teachers, parents and carers, etc.

The specific characteristics of such CVA models ought to be adapted to the ultimate purpose of the specific stakeholders they intend to inform. Other researchers have made clear distinctions between two basic types of school value-added models: a) models for school accountability and b) models for school choice. The difference between the two types lies mainly in the type of variables for which the models control. In school accountability models, further distinctions can be made as per whom the accountability is directed to; these could be either internal, when the information is used for processes that head teachers and other school authorities can carry out to analyse and improve their own practices, or external, when the information is used by government officials to assess school performance for policy-making processes.

Despite an overwhelming amount of evidence in favour of the use of such models for school accountability, government practice in Chile has been restricted to the use of school averages in the SIMCE standardised tests as indicators of school effectiveness. SIMCE stands for "Sistema de Medición de la Calidad de la Educación", which can be translated as System for the Measurement of the Quality of Education. These raw school averages in the SIMCE tests have been used indiscriminately for the purposes of school accountability and school choice, which is problematic because those two purposes entail diverse analytical choices; this is discussed in more detail in chapter 2. Nevertheless, a new accountability system that takes into account some of the principles of value-added research is being developed and is soon to be piloted in 2015 by the newly created Chilean Agency for the Quality of Education. This is certainly a valued improvement in practice; however, the main issue with this new accountability system is that it has been designed to implement Multiple Linear Regression (MLR) models to adjust school averages, which is undoubtedly a major setback. This issue is critical given the high stakes; schools performing insufficiently, as judged by this system, could face closure.

Using data from the Chilean National Pupil Database (SIMCE 2004-2006), this research demonstrates that the traditional value-added models (2-level CVA models), let alone MLR models, fall short in addressing the complex phenomenon of academic performance, especially in the context of a developing country. This is because they rather unrealistically assume, amongst other aspects, that a) there are no differences between classes within schools; b) school effectiveness does not vary across areas; c) there are no relationships between different subjects; and d) only the last school attended affects pupils' performance. Ignoring these untenable assumptions leads to a large overestimation of the school effects. The novelty of the CVA models in this research is that they extend and improve the traditional 2-level school value-added models insofar as they explicitly assess the variation between pupils, classrooms, primary schools, secondary schools and local authorities, as well as the correlation between Mathematics and Spanish Language at all levels.

External school accountability measures were derived from the extended and more complex CVA models and compared to measures derived from simpler models. This is done with the purpose of determining whether there are relevant differences and drawing conclusions with respect to the practical usefulness of the more complex models, especially concerning the implications around school accountability.

With regard to internal school accountability, further models were also specified with the purpose of analysing pupils' heterogeneity to inform school improvement processes. By definition, contextualised value-added models give insight into how the context where the learning takes place affects academic performance; however, when the focus is set on external school accountability, the models seek to control for non-malleable pupils' and schools' characteristics in order to isolate the "true" school effects. Naturally, this does not suffice from the perspective of internal school accountability, because "true" school improvement can only be brought about by intervention on critical malleable factors affecting pupils' learning. A somewhat ill-conceived school improvement can also arise from a restrictive school selectivity policy; however, results from such policies do not reflect upon real increases in school effectiveness.

A relevant source of insight into the matter of school improvement might be derived from the concept of "cultural capital" (Bourdieu, 1977), which can be understood as a set of instruments/tools that enable the appropriation of the cultural heritage of any given Society. Cultural capital is hypothesised to affect academic performance directly and indirectly through other socio-economic conditions, insofar as the value of the cultural capital possessed by individuals heavily depends on their socio-economic position.

Since cultural capital is essentially immeasurable, a latent construct approach was adopted in this research. A measure of cultural capital was constructed from a group of manifest variables related to access and use of reading materials. Then a series of multilevel models are implemented to delve into the relationship between cultural capital and pupils' academic progress in order to inform internal school accountability or improvement processes. The analyses show that the latent construct "cultural capital" has a relevant impact on analysing educational inequalities, which can eventually constitute useful information for implementing policy not only at the school level, but also at the level of local and central government.

In chapter 2, the relevant literature is reviewed and analysed in relation to the questions this research attempts to address. The overall argument of the literature review can be summarised as follows: it is known from previous research that socio-economic conditions and cultural capital affect pupils' progress and school value-added in Mathematics, Language and other subjects, as well as non-academic outcomes. However, the way in which those two elements have been analysed has not included all the relevant factors; this is especially true for the case of the analysis of school value-added, where traditional models are incomplete. On another front, socio-economic conditions and cultural capital can affect pupils' performance at different levels directly and indirectly, and these effects can even be moderated by each other and other factors and levels of aggregation. Given all these effects, it is of the utmost relevance to give careful attention to the models specified for external school accountability in order to assess schools fairly. Furthermore, internal school accountability can also be better informed via CVA models analysing pupils' heterogeneity in more depth. From a policy-making perspective, isolating schools' and pupils' effects can ultimately provide a better picture of how pupils attain what they attain, which can give us insight into how to improve overall achievement. This chapter also includes a series of definitions and historical references to situate this research in a disciplinary context and provide an overall picture of the body of knowledge of school effectiveness research and how it has been constructed over the years.

In chapter 3, the methods and data used in this research are described and explained. A special emphasis is given to the implementation of a multilevel modelling approach to analyse school value-added, since this is the main focus of this study. This chapter begins with a brief review of the MLR approach as a starting point from which more complex and realistically accurate multilevel models can be constructed, considering the complexity of the school performance phenomenon and the underlying assumptions of each subsequent model in the model building process. The chapter moves from MLR models to variance component models, where the total variance in test scores is decomposed into two sources, i.e. pupils and schools. It then moves on to the traditional value-added models, where prior attainment and socio-economic variables are taken into account, but preserving the basic 2-level structure of pupils nested within schools; then the complexity of the models increases to pair up with the complexity of the network of influences in academic achievement by specifying further levels of aggregation to take into account some of the various overlooked assumptions of the traditional CVA models. Furthermore, this chapter discusses the measurement model implemented to construct the latent variable of cultural capital. Finally, this chapter also describes the SIMCE database and the specific datasets that were used in this research and the analytical strategy to be implemented in the subsequent chapters of results.

Chapter 4 presents the first set of results focusing on the extended school value-added models for progress in Mathematics and Spanish Language. The overall purpose of this chapter is to show how a rather straightforward extension of the traditional CVA models can make a fairly large difference when using these models for school accountability purposes. The extension consists mainly of specifying two sources of variation in test scores as random effects in addition to the variation between pupils and secondary schools, i.e. the variation between classrooms within secondary schools and the variation between local authorities. The model for Mathematics progress is developed throughout the chapter to compare then the results to the model for progress in Spanish Language. These models have two main caveats: a) they are univariate and hence they convey a unidimensional notion of school effectiveness and; b) they lack control for the carry-over effects from primary schools, which makes them still liable to be deemed as incomplete. Nevertheless, their practical and substantive value lies precisely in their relative simplicity and readability for a non-statistical audience, assuming some familiarity with traditional multilevel CVA models.

Chapter 5 makes a further attempt to tackle the issues around the neglected assumptions of the traditional CVA models by analysing school value-added from a multidimensional perspective. More specifically, the implemented CVA model addresses the two main shortcomings of the univariate models for progress in Mathematics and Spanish Language presented in chapter 4; the model assesses school value-added in both subjects simultaneously while also controlling for the random effects at the levels of primary schools in addition to the levels of pupils, classrooms, secondary schools and local authorities, as already presented in chapter 4. This bivariate cross-classified CVA model proves even more effective in controlling for spurious bias in the extended univariate CVA models presented before, let alone the traditional CVA models, while it produces more precise and consistent school classifications. Nonetheless, in its strength also lies its weakness, i.e. its increased complexity, which makes it less accessible for a non-statistical audience. This chapter concludes that there is indeed a significant underlying relationship between progress in Mathematics and Spanish Language, and potentially other subjects as well, which needs to be taken into account when analysing the performance of pupils and schools.

Chapter 6 aims to analyse pupils' malleable factors affecting academic outcomes with the purpose of informing internal school accountability as opposed to the focus on external accountability in previous chapters. In this chapter, the models are set out to analyse the relationship between the latent construct of cultural capital and the progress in Mathematics and Spanish Language, preserving most of the main features of the CVA models in previous chapters, namely: the relationship between both subjects is taken into account; there is a reasonably complex model structure of pupils nested within secondary schools within local authorities and; there are reasonable controls for socio-economic and prior attainment in order to provide sufficiently contextualised information. As mentioned before, the construct of cultural capital proved useful to analyse malleable pupils' heterogeneity in Spanish Language progress as well as in Mathematics and consequently to inform school improvement processes by showing the potential impact of reading habits and access to reading materials on academic outcomes.

Finally, chapter 7 presents the overall discussion of the results from the statistical analyses of previous chapters. The discussion contrasts the empirical results with the underlying theory around four main topics that work together as a way to introduce the conclusions of this thesis that are presented in chapter 8. The conclusions revisit the research questions spelt out in chapter 2 and develop three main ideas around them. Additionally, the implications for public policy are discussed in close link to these three main conclusions throughout chapter 8. Lastly, some recommendations for further research on school value-added, school accountability and other specific topics are outlined.

Chapter 2: Literature Review

2.1. Introduction

In this chapter, the relevant literature in the research areas related to school value-added and school effectiveness is reviewed to provide the theoretical background that helps unfold and guide the research questions at the end of the chapter. The review focuses on topics such as external and internal school accountability, school choice, the methodological challenges of analysing school value-added along with its purpose and the relationship between the socio-economic characteristics of the pupils and the socio-economic context of the schools, their academic performance and how schools contribute to this process. In order to contextualise this research and its purposes, the chapter also presents and discusses the current state of the art of Chilean-based research on school effectiveness along with some background information on the education system as well as the current government practice regarding school accountability.

This chapter deals with five main topics. Section 2.2 aims to situate this research in a disciplinary context and to give insight into how models of educational effectiveness and school value-added came about, focusing on the history and definition of the concepts of school effectiveness, school improvement, educational quality and how these are interrelated. In section 2.3, the aim is to discuss the most relevant theoretical, empirical and methodological aspects of school value-added models, in order to establish the foundations on which more precise and fairer internal and external school accountability systems can be built. Next, in section 2.4, the purpose is to discuss the concept of school accountability and its implications as this is central to the practical purpose of this research. Section 2.5 builds upon the concepts and issues around school value-added research discussed in the previous section to shift towards the analysis of internal school accountability; this is done by outlining the way in which structural socio-economic factors operate in conjunction with cultural capital to shape pupils' academic success. Section 2.6 seeks to describe the context in which the Chilean education system is embedded, focusing on education inequalities due to socio-economic background. Finally, section 2.7 summarises the key concepts and section 2.8 enunciates the research questions to be tackled in this thesis and the rationale behind them.

2.2. The foundations and history of the study of educational quality

The concept of school effectiveness came along with the notion of quality of education. It was initially developed during the period after the Second World War, when educationists saw the potential of the applications of the concept of quality that was firstly used in industrial contexts for the control of the quality of products and services (Kumar & Sarangapani, 2004; Sallis, 2002). Afterwards, a rather managerial perspective was introduced into the educational field, but the concept of quality grew apart from that approach and now has its own framework with substantive meaning and applications. This has brought about a more educationally-focused body of research on what it means to have a good quality school (or school system), how to measure quality and its implications. However, as it will be presented later on, the implementation of the idea in the reality of schools by different actors can vary widely, with the intervention and inclusion of different elements involved in the school (or school system) processes.

School effectiveness research has also a vast literature, which can be traced back to the decade of the 1960's, with the appearance of the first studies which claimed to be assessing the level of effectiveness of schools (Coleman et al., 1966; Jencks et al., 1972). From this starting point, numerous different perspectives have arisen, as it will be discussed in later sections (2.2.2 and 2.2.3).

More recently, public and international academic debate on education has focused mostly on policy issues surrounding external school accountability and raising academic standards. Kumar and Sarangapani (2004) argue that a special emphasis on this has been put in developing countries; however, their approach has been rather technocratic. The authors go further and claim that the ill result is that developing countries have ended up in an unintended and endless "catching up" race with the school systems of developed countries, instead of creating sustainable conditions for the improvement of education for all children.

School effectiveness research (SER) is strongly related to the development of the research and policy on educational quality. Luyten et al. (2005, p. 249) define SER as "the line of research that investigates performance differences between and within schools, as well as the malleable factors that enhance school performance". This line of research has a great deal of diversity and perspectives; some authors (see for example: Creemers, 2007; Luyten et al., 2005; Reynolds et al., 2000; Scheerens, 2000) have developed meta-analyses of SER sorting out the studies based on a number of different criteria, such as purpose, scope, methodologies involved, etc. Considering their object of study, a number of major strands of SER can be defined. The first major strand to be taken into account is that of the 'school effects research' (Reynolds et al., 2000), which includes studies concerned with the properties of the school effects, i.e. the outcomes produced by schools at the level of pupils (results in standardised tests), with analyses ranging from the adaptation of input-output models in its early stages to the implementation of multilevel models in the most recent studies. Regarding the analysis of processes within schools, there is another major strand in SER, defined by Reynolds et al.

(2000) as ‘effective schools research’. This strand is focused on the analysis of the internal processes of effective schools, which have developed from case studies of outlier schools, to the most recent focus on the analysis of classroom within schools. Furthermore, there is an increasing interest in studying ineffective schools, as effective schools have been more frequently researched.

Another important strand in school effectiveness research is the one that analyses the processes by which educational institutions can be changed to improve their results. This is known as ‘school improvement research’, which can be understood as the application of the generic school effectiveness models to practical situations in real schools. This line of research is concerned with the educational policy implemented at all levels of the system, i.e. school districts, institutions and classroom improvement programmes.

From another conceptualisation, Scheerens (2004, p. 2) includes in this classification of the school effectiveness research areas, the "research on equality of opportunities in education and the significance of the school in this", which may be understood as a line of studies concerned with the sociological analysis of educational inequalities. This area of research might also be categorised as a subset of the "school improvement research", in terms of its contribution to the understanding of the role that schools play in society and the way to improve educational outcomes. This author further defines the "economic studies on education production functions", as studies concerned with the analysis of the economical processes involved in the educational systems. In terms of public policy, the author also includes the evaluation of compensatory programmes in this classification, but these studies can be considered as a part of the school improvement strand, given that they analyse the processes involved in (improvement) programmes implemented in schools.

In spite of the definitions presented above, this review will focus on a more generic and summarised categorisation of lines in SER, which has on the one hand the ‘School Effectiveness’, i.e. the studies concerned with the analysis of the school, including effects of schooling, processes, student outcomes, from a mainly quantitative point of view, and on the other hand the ‘School Improvement’ strand, which is concerned with the implementation of school policy interventions towards improvement in the provision of schooling. This implies a more marked focus on institutional outcomes, conditions and organisation, which are very frequently analysed from a qualitative point of view. In this study, these two lines of thought are considered to be complementary as one is concerned with the description of the current conditions and the other takes into account what has been analysed in order to improve schools (or the school system).

2.2.1. A brief historical overview of school effectiveness research

In the very early stages of development of school effectiveness research, which extended from the mid 1960's until the early 1970's, the predominant paradigm was the 'input-output model'. These studies were mostly economically driven, focusing on inputs such as school resources and student background. The first conclusions that emerged from this line of research indicated that "school made no difference" (Coleman et al., 1966; Jencks et al., 1972). Such conclusions were drawn upon the evidence found that the differences recorded in pupils' achievement were more associated with socio-economic background rather than school conditions (Reynolds et al., 2000).

These initial conclusions about the role that school played in the learning and attainment of students were heavily criticised from the early 1970's. Right after the first input-output model studies, a number of studies were conducted to dispute their results. In these studies, the notion of 'educational process' was introduced and therefore variables such as library resources, classroom information, school assignments, strong leadership, high expectations, etc. were included in the school effectiveness models. The main difference between this particular line of studies and its predecessors is the dynamism added by the analysis of the ongoing school processes (Reynolds et al., 2000), which were analysed in early studies as static input school characteristics.

From the late 1980's, school effectiveness research has developed greatly, mainly because of the development of highly sophisticated statistical methods such as multilevel modelling (Hill & Rowe, 1996; Reynolds et al., 2000). A more thorough account of recent developments is discussed in later sections, more specifically in section 2.3 on school value-added research.

2.2.2. The quality of schooling and school improvement

The concept of "quality" emerged in industrial contexts after the Second World War, and was associated with the quality in production processes. The United Kingdom and the United States only began to be attracted to this concept in the 1980's after the enormous success of Japanese industries during the 1970's (Sallis, 2002). At that time, managerial sets of tools, techniques and norms were created to improve, control and ensure the quality of products and services; these include, for instance: ISO9000 (International Organization for Standardization, 2015), the EFQM model (European Foundation for Quality Management, 2015), the Deming Model (Deming, 1993), amongst others.

During the 1980's and 1990's, these models started to be implemented in educational institutions (Hides et al., 2004; Izadi et al., 1996; Lundquist, 1997; Michael et al., 1997; Van

Den Berghe, 1998). However, this perspective has been heavily criticised for being too inclined to managerial practices, such as defining students as customers or clients (Cuthbert, 2010; Emery et al., 2003; Franz, 1998; Kumar & Sarangapani, 2004; Sax, 2004).

As a response, researchers in the field of Education advocated for a vision on quality based on the knowledge gained by means of educational studies (Emery et al., 2003; Scheerens, 2000). Thus, school improvement research emerged during the late 1970's. The main focus of this line of research was on fostering effective schools rather than describing them as opposed to school effectiveness research (Reynolds et al., 2000). The ultimate goal was to move towards greater equity in schools by implementing improvement programmes focused on fostering the academic progress of disadvantaged pupils.

One of the first and more widely known of these school improvement models is the "five factor model of effectiveness" (Edmonds, 1979). This model included the following factors: the leadership of the directive staff, an emphasis on instruction, a safe and orderly organisational climate, the expectation of teachers, and the permanent monitoring of pupils' achievement and progress. This particular model was created from the work of the practitioner community and gained wide acceptance amongst educational practitioners during the 1980's (Reynolds, 1995), but not without heavy criticism (See for example: Coe & Fitz-Gibbon, 1998; Creemers, 1994) because of its alleged lack of theoretical grounding. Later on, other models including additional relevant factors were proposed; an example of this is the "eleven factor model" (Stoll & Mortimore, 1997), which emphasised the relevant influence of the school vision and goals, pupils' discipline, rights and responsibilities, staff development, and parental involvement on pupils' progress.

The growing concern with the problem of inequalities in education systems, especially in developing countries, has led international organisations such as the United Nations Children's Fund (UNICEF) and the United Nations Educational, Scientific and Cultural Organization (UNESCO), to develop their own frameworks to assess the quality of education, following the conceptualisation and general precepts of school improvement research.

For instance, UNICEF (2000) highlighted the main aspects of educational quality that need to be taken into account from the wider perspective of child protection in the context of developing countries. This institution described the need to ensure: "a) learners who are healthy, well-nourished and ready to participate and learn, and supported in learning by their families and communities; b) environments that are healthy, safe, protective and gender-sensitive, and provide adequate resources and facilities; c) content that is reflected in relevant curricula and materials for the acquisition of basic skills, especially in the areas of literacy, numeracy and skills for life, and knowledge in such areas as gender, health, nutrition, and peace; d) processes through which trained teachers use child-centred teaching approaches in well-managed classrooms and schools, and skilful assessment to facilitate learning and reduce disparities, and finally; e) outcomes that encompass knowledge, skills and attitudes, and are linked to national goals for education and positive participation in society" (p. 3).

On another front, UNESCO (2004) attempted to define a comprehensive framework for understanding education quality from a system-wide perspective. According to this institution, educational quality needs to be analysed as a country-wide series of processes, which can be classified into four broad areas: a) learner characteristics, which include aptitudes, perseverance, school readiness, prior knowledge, barriers to learning; b) enabling inputs, which includes teaching and learning (time, methods, assessment, feedback, incentives and class size), materials, human resources and school governance; c) outcomes, which should at least include literacy, numeracy, life skills, creative and emotional skills, values, social benefits, and finally; d) context, which involves the analysis of the influence of economic and labour market conditions in the community, socio-cultural and religious factors, educational knowledge and support infrastructure, public resources available for education, competitiveness of the teaching profession on the labour market, national governance and management strategies, parental support, national standards, public expectations and so forth.

As can be seen from the foregoing, the issue of educational quality is complex and multifaceted. Nonetheless, for the sake of simplicity, educational quality can be thought of as having two main foci, namely: pupils' learning outcomes and school processes. These two foci are by no means incompatible; on the contrary, they are at the very least complementary. On the one hand, focusing on learning outcomes implies driving attention to externally observed results, which is the recording of the actual school effects on pupils, controlling for exogenous and endogenous causes. This has the advantage of having the capability of generalising and predicting pupils' performance or learning outcomes, either in the future or for other contexts. This perspective is usually known as "School Effectiveness Research". On the other hand, focusing on school processes implies driving the attention to the quality of the school provision by describing or analysing the school conditions, infrastructure, the qualifications of the staff, the school management style, amongst other variables of interest. The advantage of this viewpoint is its in-depth knowledge of the schools. This perspective is usually known as "School Improvement Research".

Regardless of whether the focus of the research is set on academic outcomes or school processes, any educational study needs to incorporate the analysis of structural influential factors external to the teaching and learning process, such as the socio-economic and demographic background. Controlling for such influential factors can serve basically two purposes: firstly, it allows holding schools accountable for what they are "truly" responsible, and secondly, it allows focusing interventions on the factors that schools can reasonably modify. This will be discussed more deeply in sections 2.3, 2.4 and 2.5.

2.2.3. The main contributions of school effectiveness research

School effectiveness research has now an important place in educational research, mainly because it has contributed throughout the years in a number of crucial aspects in terms of research, public policy and teaching practice. As noted earlier, during the 1960's educational research had created a strong belief that "schools do not matter", and therefore, pupils' educational and occupational future were determined by their socio-economic background (Coleman et al., 1966; Jencks et al., 1972). This conclusion turned out to be potentially very harmful in terms of public policy, as it implied that there seemed to be no point in making large school investments from the State.

Not only school effectiveness research has contributed to the dismissal of the "schools do not matter" notion, but it has also been able to identify a number of key school factors that produce measurable positive results (Reynolds, 1995). Moreover, this research line has contributed to the belief that "teachers matter" as well and, thusly, created a sense of responsibility and professional acknowledgement of the teaching career. In accordance with all the contributions mentioned, researchers have created a robust body of knowledge "known to be valid" (Reynolds, 1995), a foundational understanding of "what works", and a wide range of tools to diagnose schools and provide training for practitioners, as well as policy-makers.

Despite all the contributions, widespread diffusion of school effectiveness research results has been misinterpreted and wrongfully used to create a belief "that schools do not just make a difference, but that they make all the difference (…)" (Reynolds, 1995, p. 59). This belief is especially useful for policymakers to promote educational compensatory programmes and specific school interventions; however, from a sociological point of view, it is not possible for schools to compensate for inequities in society, because the very fact that schools exist can facilitate, foster or reproduce social inequalities. This will be discussed more thoroughly in section 2.5.

More recent contributions of this research strand are related to the sophistication of the models to make fairer school comparisons, holding schools accountable for the circumstances over which they have control and inform school improvement interventions in a better way

(See for example: Leckie, 2009; Timmermans et al., 2013). A more thorough account of such contributions is unfolded in later sections, especially the section on school-value added research (2.3).

2.2.4. Criticisms to school effectiveness research

Undoubtedly, school effectiveness research has made considerable contributions to the educational field, which are almost paired with the amount of criticism it has raised throughout the years. The most relevant sources of criticism to this line of research can be found on three key aspects: its political-ideological nature, its theoretical limitations and its research methodology.

Regarding its political-ideological nature, Luyten et al. (2005) argue that school effectiveness research lacks ideological independence, which is reflected in the fact that the issues around the appropriateness of national goals and its correspondence to standardised test are hardly ever addressed by these studies. Furthermore, these authors criticise the objectivity of such studies, because of the little attention paid to the extent to which schools can boost children's achievement. In other words, the actual size of the school effects is not a clear matter and this is highly relevant in terms of public policy, i.e. to what extent can children overcome social inequities by means of education? The fuzziness of the response to this question may be leading to a school inspection and culture based on blame (Rea & Weiner, 1998) and guilt (Hargreaves, 1994). This is also related to the high-stakes testing approach (Au, 2007), which is discussed later in this review.

Likewise, the theoretical limitations of SER have to do with the stronger preference for statistical grounding to justify the inclusion of variables in the models, rather than theoretical grounding (Coe & Fitz-Gibbon, 1998). In addition to this, some criticisms come from some authors identifying blind spots such as: the stronger emphasis put on the most malleable factors as a result of the goal of SER, that is identifying the factors that lead to the best results (Scheerens & Bosker, 1997). Similarly, according to Luyten et al. (2005), authors have tended to overlook the question of which school and classroom structures are the best for different types of students, i.e. there is no reliable answer to the question of differential effects. Moreover, the effectiveness criteria sets have been criticised for their narrowness, which occurs when researchers adjust their theoretical framework to the data available (Coe & Fitz-Gibbon, 1998), leading SER to be excessively oriented towards accountability (Lingard et al., 1998). This is why various researchers (Au, 2007; Sahlberg, 2007, 2010) advocate for a new approach on accountability that encourages school improvement instead of focusing on rewards and sanctions. This is discussed in more detail later in this review (section 2.4).

The criticism to the methodology used in SER has to do firstly with the definition of school effects. Coe and Fitz-Gibbon (1998) remark that what researchers most commonly define as the "school effect" is the variance in the data that has not been explained by the school characteristics (input) controlled in the study. However, this conceptualisation of school effects is widely accepted and follows rigorous methodological and statistical criteria (See for example: Goldstein, 1997; Raudenbush, 2004; Timmermans et al., 2011); this is discussed more deeply in the section of school value-added research of this review (section 2.3). Secondly, other methodological issues are related to the overwhelming attention drawn to effectiveness and the scarce importance given to the study of ineffectiveness. In this respect, Reynolds and Teddlie (2001) argue that SER has tended to focus excessively towards normal or average schools. Likewise Scheerens and Bosker (1997) advocate for research delving into the processes by which ineffective schools improve and effective schools decay, by combining the utilisation of large-scale national databases to examine school development deeply, and more rigorous in-depth data collection.

The next section integrates the foregoing foundational concepts into the most recent discussions around school effectiveness and school value-added, which sustain this research.

2.3. School value-added: foundations, methodological challenges and extensions

Educational inequalities have been a great concern for practitioners, policy-makers and researchers during the last decades. Views vary between those authors who claim that "schools do not matter" (see for example: Coleman et al., 1966; Jencks et al., 1972), to those who claim that in fact "schools and teachers do make a difference" (or can add value to) in the educational trajectories of disadvantaged pupils (See for example: Reynolds, 1995; Scheerens & Bosker, 1997; Scheerens, 2000, 2004). The extent to which schools can contribute to overcome socio-economic inequalities has been one of the main questions that educational researchers have tried to address; the question of how effective schools are, from the perspective of disadvantaged pupils, refers to how schools foster pupils' performance beyond what is expected given their initial attainment and the obstacles they have to overcome from a given socio-economic situation (Creemers, 1994; Goldstein, 1997; Reynolds, 1995; Scheerens & Bosker, 1997). In other words, the main "effectiveness question" from the perspective of this research field is "Which schools actually add 'value' to the educational trajectories of their pupils?"

The discussion about "value-added" has its origins in the field of measuring school effectiveness, and simply refers to the amount of value added to any given student's test outcome that is uniquely attributable to attending a particular school (Scheerens, 2000, pp.

18–19). School effectiveness itself is a controversial term. According to Goldstein (1997, 2001) value-added indicators are inherently uncertain and are relative to other schools' performance. Thus the most appropriate concept is "relative school effectiveness". An effective school is defined here as one that encourages the progress of all its pupils beyond what is expected given their initial levels of achievement and socio-economic (and cultural) background, ensuring in a sustainable manner, that all students get the highest possible outcomes and improvement in every aspect of their development and performance (Scheerens, 2000). Value-added is "an indication of the extent to which any given school has fostered the progress of all students in a range of subjects during a particular time period in comparison to the effects of other schools in the same sample" (Sammons et al., 1997, p. 24).

From this discussion, it is also derived that the terms "school effectiveness" and "school value-added" can be used interchangeably, insofar as, underlying both concepts, there is the idea of measuring the contribution of schools to the progress made by pupils in diverse subjects.

2.3.1. The measurement of school value-added and its practical implications

One indicator of school performance has been the averages in standardised tests. This has proved to be flawed given the risk of drawing seriously misleading conclusions as a result of the "ecological fallacy"[1], since relationships between variables might well be radically different at the school level compared to the pupil level, which does not allow us to infer about differential effects when comparing groups of students within schools (Gibson & Asthana, 1998; Goldstein & Spiegelhalter, 1996; Goldstein, 1997, 2001). The problem is more pronounced when yearly comparisons of school averages are made, since these comparisons are performed between different cohorts and without controlling for pre-existing differences between pupils (Goldstein, 1997). For instance, a sudden rise in the average of a particular school could be due to an uncontrolled increase in the number of better-performing pupils who are admitted to the school instead of an actual improvement of the school's effectiveness. In other words, schools could appear to have improved their scores simply by recruiting better-performing pupils. Despite all these downsides, current government practice in Chile still employs school averages as indicators of effectiveness (San Martín & Carrasco, 2012). Although this practice is changing as it will be discussed later on.

Researchers have advocated for a change in these practices, because of the indisputable amount of evidence in favour of more sophisticated statistical techniques for the analysis of school performance for public accountability (Goldstein & Thomas, 1996), such as multilevel

[1] The ecological fallacy (Robinson, 1950) is a mistaken conclusion about the individuals of a group drawn from the sole analysis of group (aggregated) data.

modelling, which has become standard practice in educational research. These models traditionally analyse variation in test scores at the level of pupils as well as schools, and have been proved to be fruitful.

There is a variety of multilevel models that have been used to analyse school effectiveness. For a start, "gross school effects" or "type 0" value-added models (Timmermans et al., 2011) specify a certain standard measure of attainment as the outcome and account only for the variance at the school and pupil level. They are empty models as described in the multilevel modelling literature (Goldstein, 2011). Nevertheless, it is not accurate to regard such models as value-added, because they do not measure progress. The simplest school value-added model that can be specified is the "raw" value-added (VA), which measures the progress made by the students controlling for their prior attainment. This model is said to measure the contribution of the school (school value-added) to the change in pupils' attainment (progress), but does not consider contextual information and thus it is deemed to be a "raw" measure of value-added. This definition corresponds to various names by various experts: “value-added” was previously used by the English Department for Education (Ray, 2006), "type AA" value-added (Timmermans et al., 2011) and "adjusted (school) comparisons" (Goldstein, 1997) are also amongst the "synonyms".

From this (VA) model, it is possible to derive raw school value-added estimates, which correspond to the shrunken residual point estimates (also known as empirical Bayes residuals) of the multilevel model, after controlling for prior attainment (Goldstein, 1997; Ray, 2006). The problem with this approach is that school prior attainment averages appear highly correlated with the value-added scores, which results from misspecification of the raw value-added models (Foley & Goldstein, 2012; Goldstein et al., 2007). This misspecification implies that the school effectiveness indicators are biased towards those schools serving the highest prior-attaining students. Consequently, it is necessary to account for differential effects in the models (Foley & Goldstein, 2012). In other words, a (raw) VA model is misspecified insofar as it fails to control for relevant factors that influence academic performance while it also fails to meet relevant statistical assumptions.

Alternatively, contextualised value-added (CVA) models control for prior attainment in the subject of interest, as well as variables related to pupil socio-economic and demographic characteristics and school context (many authors agree on this concept, see for example: Foley & Goldstein, 2012; Goldstein, 1997; Ray, 2006; Sammons et al., 1993). In other references these models come under the names "type A", when controlling for student-level covariates only (Raudenbush, 2004; Timmermans et al., 2011), "type B", when controlling for pupil-level covariates and compositional school-level variables (Raudenbush, 2004; Timmermans et al., 2011), and "type X", when controlling for pupil-level covariates and compositional school-level variables and other non-malleable school factors (Timmermans et al., 2011). This was the approach and concept adopted by the English Department for Education until 2011 (Ray, 2006).

Beyond the statistical superiority of this approach as mentioned above (i.e. resolving the bias due to misspecification) there have also been policy relevant arguments in favour of CVA scores estimation. In particular, educational research has demonstrated throughout the years that the main sources of differential effects in school performance are unsurprisingly associated with socio-economic and demographic characteristics of the pupils and the schools. With respect to public policy, the use of these variables is relevant since using only school averages (without controlling for socio-economic variables) as indicators of effectiveness implicitly makes the schools accountable, for better or worse, for circumstances they cannot modify or for which they are not responsible (Ray, 2006).

However, CVA scores (and any other VA evaluation system) are not error-free, because they are products of statistical modelling processes, and hence, inherently uncertain, which is why confidence intervals should be and have been routinely estimated and published in performance tables in England (Foley & Goldstein, 2012; Ray, 2006). The use of CVA scores and their corresponding confidence intervals has some limitations related to high levels of uncertainty, especially when making predictions about future school performance (Leckie & Goldstein, 2009, 2011b). Despite these limitations, Leckie and Goldstein (2011b) point out that performance tables are still retrospectively informative and provide highly relevant information for holding schools accountable.

Another aspect of the measurement of value added is model complexity. Customarily, these value-added studies implement multilevel models where variation in standardised test scores comes from only two sources: schools and pupils. One of the main underlying assumptions of these models is that school value-added is a latent trait that can be estimated from the analysis of the performance of the pupils within schools (Timmermans et al., 2013). Some other relevant and perhaps overlooked assumptions are: a) the last school a pupil has attended concentrates the whole amount of value-added throughout their educational trajectory; b) there are no differences between classrooms within schools; c) schools located in diverse geographical or administrative areas do not differ significantly; and d) there is no relationship between subjects (this is discussed in more detail in section 2.3.4). These assumptions are not only strong but rather unrealistic when analysing school performance. In the next section, these assumptions are discussed and debunked.

More recent research has pointed out the necessity of increasing the complexity of the school value-added models by specifying additional levels of variation, which can be either completely nested or follow a cross-classification pattern (De Fraine et al., 2003; Goldstein et al., 2007; Leckie, 2009; Martínez, 2012; Rasbash et al., 2010; Timmermans et al., 2013). In these studies, the levels of neighbourhoods, classrooms, primary and secondary schools, as well as local education authorities were specified (although not all the levels simultaneously in all studies) and they were all found to be relevant sources of variation in pupils' test scores. However, these studies are based in the United Kingdom, the United States and the Netherlands, which are all developed countries whose education systems are embedded in radically different socio-economic contexts compared to the Chilean education system.

As Scheerens (2000) points out, international research has demonstrated that the size of the school effects in developing countries far exceeds the size of the school effects in developed countries. This is because the education systems of developing countries have to deal with higher levels of deprivation and multiple inequalities. This would potentially make plausible the existence of larger effects proceeding from factors external to the school and pupils, which renders research based on national contexts such as the Chilean even more relevant, especially for its potential impact on policy-making.

Chile's inequality is reflected in the Gini coefficient[2] of 52.1 (The World Bank, 2013) and the loss of 19% in its Human Development Index (HDI)[3] due to inequality (United Nations Development Programme, 2013). Large HDI differences by region and between localities can also be found (United Nations Development Programme, 2004). This makes plausible the existence of significant differences in pupils' achievement according to the geographical location of their school. On another front, Chilean schools have historically put in place streaming practices, where pupils are assigned to different classrooms according to their ability level (Torche, 2005), which makes those schools likely to reveal differences in value-added between classrooms. Despite the evidence in favour of specifying value-added models that are more complex than the traditional 2-level model, only two recent studies conducted in Chile (Manzi et al., 2014; Mizala & Torche, 2012) have addressed issues related to the effects of streaming, the geographical location of the schools or the carry-over effects from primary school. This will be discussed in subsequent sections (2.3.2 and 2.3.3).

[2] 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 (2011).
[3] The Human Development index (HDI) is a summary measure of achievements in key areas of Human Development, namely: Health, Education and Income. For a more thorough explanation, see: (Anand & Sen, 1994).

2.3.2. The assumptions underlying traditional school value-added models

Traditionally, school effectiveness models have been employed to assess the variation in pupils' scores sitting one or more standardised tests coming mainly from two sources: the pupils and the schools. This is done with the objective of distinguishing between the effects of pupils' abilities and characteristics from the effects of the characteristics of the schools pupils are attending at the time of the tests. Ultimately, one of the most important aims of such models is to ascertain the impact of school policy on pupils' progress. School policy in these models is assumed to be related to all those effects that arise as unexplained variance in a model, after controlling for relevant non-malleable school and pupils' characteristics. This is achieved by implementing a standard two-level model, where pupils are nested within schools.

Recent research (Goldstein et al., 2007; Leckie, 2009; Rasbash et al., 2010) has found that the somewhat standard approach to analyse school effectiveness is incomplete insofar as there are assumptions underlying these models that are simply untenable for statistical and, more importantly, substantive reasons. Although making assumptions is an essential and inevitable part of any statistical modelling process, the consequences of these assumptions and their suitability for the purposes of the research must also be assessed. This research may indeed overlook certain aspects of the school performance phenomenon and make hard-to-meet assumptions due to omission of variables. Nevertheless, the results of this research certainly constitute valuable information for implementing improvements to current practice in Chile. The value of the results lie precisely on moving towards a more transparent analysis of the substantive implications of the most frequently unquestioned assumptions in school effectiveness research.

Firstly, schools may put streaming in practice, as is the case of Chile (Torche, 2005), which makes differences between classrooms (within schools) highly likely. The significance of the classroom effects has been demonstrated by various authors (See for example: Cervini, 2009a; 2012; Murillo & Román, 2011).

Secondly, assuming that the last school a pupil has attended at the time of the standardised tests is the only one affecting their academic performance is untenable, because pupils often have attended other schools besides that one. Moreover, in the case of standardised tests taken during secondary schooling, it is highly relevant to differentiate pupils according to the primary schools they attended in order to estimate the long-term carry-over effects that might still be affecting subsequent attainment. This second point has been deeply explored and its relevance demonstrated by various authors (Goldstein et al., 2007; Hill & Goldstein, 1998; Leckie et al., 2010).

Thirdly, the standard two-level model approach to school effectiveness also fails to discern geographical effects. When no other higher-level (above the school level) is specified, it is assumed that there is no other source of variation beyond the differences between schools. In doing so, the differences between areas are disregarded as trivial and geographical inequalities that might affect school effectiveness become obscure and confounded with the school-level effects. In other words, if pupils in a particular school are more likely to obtain similar results due to having shared a school environment throughout the years, then schools in a particular geographical area are also more likely to have similar levels of effectiveness due to sharing a geographical setting as compared to schools in other areas. This school geographical clustering has been shown to be statistically significant and substantively relevant in other studies (Cervini, 2009a; Leckie, 2009; Plewis, 2011; Rasbash et al., 2010).

2.3.3. The practical implications of the traditional school value-added models

The previously mentioned adjustments of the traditional approach to school effectiveness are also of the utmost importance from the perspective of diverse stakeholders. Firstly, adjusting for classroom effects, from the perspective of the school administration, allows focusing school policy interventions on particular aspects that may be affecting classes differentially. On the other hand, from the school choice point of view, parents and carers are generally unable to choose the class to which they would wish their pupils to be assigned. Hence, an overall measure of school effectiveness adjusted for classroom effects is more precise with regard to within-school differences.

Adjusting for effects from previously attended schools has implications for school administration and school choice as well (Goldstein et al., 2007; Leckie, 2009). Having information of the primary schools from which pupils come allows schools, for instance, to pursue remedial actions for disadvantaged pupils. From the parents’ and carers' perspective, this has two main implications: a) school choice is based on school accountability information that is less obscure and distinct from carry-over effects from other schools; and b) an estimation of the effect of pupil's mobility can be derived.

Moreover, adjusting for geographical effects is useful, for instance, for school accountability purposes insofar as these adjustments allow policy-makers and government officials to have geographically contextualised information about the schools, which is necessary to focus the use of resources in a more efficient way. On another front, this information is essential for policy interventions at the level of local governments, simply because they are limited to intervene on a circumscribed territory. Last, but by no means least, geographically contextualised information about school effectiveness is crucial for school choice, since parents and carers are most often limited to choose only from those schools located in the area in which they reside.

2.3.4. School value-added as a multidimensional phenomenon

On another front, traditional school value-added models assume that the subjects under assessment are unrelated to each other as mentioned in section 2.3.1. This is because even though a school accountability system may include information on several subjects, as is the case of the performance tables published by the England and Wales Department for Education (Ray, 2006), the underlying statistical models do not take the relationships between them into account. Assuming that there is no relationship between subjects has basically two implications: a) at the pupil level, a null relationship between two subjects, say, Mathematics and Language, would imply that pupils have learned both subjects completely separated from each other; and b) at the school level, this would imply that subjects are taught in an isolated way.

Several researchers have found that there is a significant link between language development and mathematical attainment. On the one hand, there is research on how linguistic impairments can affect mathematical skills development during childhood (Cowan et al., 2005; Donlan et al., 2007; Simmons & Singleton, 2008); while, on the other hand, there is research on how generic and specific language skills can affect the development of mathematical skills of pupils in general (Hecht et al., 2001; Vukovic & Lesaux, 2013).

From the perspective of the linguistic impairments, Donlan et al. (2007) found that children with a specific language impairment performed significantly lower in a variety of mathematical tasks than the children in the control group. More specifically, these authors (Donlan et al., 2007) concluded that language impairment is an inhibitor factor when it comes to the development of mathematical skills, such as the acquisition of the number sequence, the development of calculation skills and the acquisition of the place-value principle; while this is not the case for learning arithmetic principles, which seems to be unaffected. Cowan et al. (2005) also found evidence that language comprehension affects mathematical skills, specifically tasks related to addition combinations and relative magnitude. Finally, Simmons and Singleton (2008), through meta-analysis, found that the difficulties endured by dyslexic pupils are not limited to language attainment, but are consistently related to mathematics performance as well.

From a wider perspective, other authors also found evidence of this underlying relationship between language abilities and mathematics performance of pupils in general. Hecht et al.

(2001) reported that general verbal ability is persistently correlated with mathematics computation skills in pupils from Year 2 to Year 5. Along these lines, Vukovic and Lesaux (2013), using mediation analysis on a sample of Year 3 pupils, also found that verbal analogies had an indirect relationship with arithmetic knowledge via symbolic numeric skills, while phonological decoding was directly associated with arithmetic performance.

More recent research conducted in Italy by Sulis and Porcu (2014) provided further insight into the relationship between Mathematics and Language attainment at multiple levels, more specifically at the level of schools and geographical areas. The authors found evidence that schools that perform well in Mathematics also perform well in Language. One limitation of this study is that no control for prior attainment was specified, and hence, neither pupils' progress nor school effectiveness could be measured. This correlation between subjects at the school level seems to be pointing out that school performance is multifaceted. Thus, school effectiveness models need to incorporate information on a variety of educational outcomes to be able to ascertain these underlying correlations.

2.3.5. School value-added research in Chile

In Chile, proper school value-added research has only been made possible recently. The SIMCE evaluation system, until 2006, considered only once the assessment of pupil's achievement of the National Curriculum goals throughout their educational trajectories. From 2006 onwards, pupils would sit the SIMCE tests either in years 4 and 8; years 4 and 10; or years 8 and 10. Only from that point onwards, could school effectiveness studies truly control for prior attainment in the evaluations and hence school value-added studies were made possible.

Very few Chilean-based studies have discussed some of the issues around analysing the variation in progress coming only from pupils and schools in school effectiveness models. Mizala and Torche (2012) implemented two univariate cross-sectional (without controlling for prior attainment) 2-level models for attainment in Mathematics and Language of pupils in Year 4 in 2002 and pupils in Year 8 in 2004. 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 socio-economic stratification. The authors also attempted to control for within-school diversity by adding the standard deviation of an indicator of family socio-economic status (SES), which they produced via latent factor analysis, as an explanatory variable at the school level. Regarding geographical location of the schools, the authors controlled for whether the schools were located in rural areas; however, this indicator variable does not account for other presumable regional differences, for instance between large cities and small towns.

Another recent study by Manzi et al. (2014) proposed an alternative school value-added model to control for endogeneity present in the univariate value-added model for progress in Mathematics resulting from a 'traditional approach', i.e. a 2-level model of pupils nested within schools with random intercepts only, where prior attainment scores are correlated with the school value-added estimates, thus violating one of the assumptions of the multilevel modelling approach. The authors implemented a 2-level random coefficients contextualised value-added model controlling for parental qualifications, prior attainment, average school prior attainment and socio-economic status. They demonstrated how the inclusion of average school prior attainment controls for bias in the estimation of the school effects induced by school selectivity. Nevertheless, these authors made no specific reference to within-school or between-localities effects.

Although there are no studies in Chile that analyse explicitly within-school or between-locality effects, in neighbouring Argentina, Cervini (2009a, 2009b) demonstrated that classroom and geographical effects are indeed relevant. This author found that the between-school variation in Mathematics scores was greatly overestimated when using the traditional 2-level model in comparison to a 5-level model of pupils nested within classrooms, schools, municipalities and states. These studies are certainly relevant considering the similarities regarding the context in which the Chilean and the Argentinian education systems are embedded.

2.4. School value-added models as a research tool for school accountability

One of the most important purposes of analysing the contribution of schools to the academic progress of pupils is school accountability. School accountability in its simplest form can be understood as the identification of the responsibility that lies with the schools with respect to their pupils' learning. The traditional approach to implement such notions conveys the definition of performance standards, school monitoring and inspection, pupils' achievement testing, as well as rewards and sanctions according to performance (Sahlberg, 2007, 2010). Furthermore, as in the case of England and Wales, the accountability system can also involve the publication of school performance tables for public scrutiny (Ray, 2006). This widely-known and traditional approach can be referred to as external school accountability (San Martín & Carrasco, 2013). Some authors (Au, 2007; Sahlberg, 2007, 2010) also refer to this approach as high-stakes testing, where the accent of the accountability system is on the consequences for schools.

Nonetheless, the purpose of a school accountability system is not necessarily to inform external stakeholders. Information emerging from such systems can also be used by the schools to improve their own practices. This is known as internal school accountability (San

Martín & Carrasco, 2013). Along these lines, Sahlberg (2007, 2010) advocates for an intelligent accountability approach, where the main focus is set on the information that is useful to help schools improve their practices and hence their results, while satisfying the demands for holding schools accountable externally.

In this research, it is shown how school value-added models can serve both the main purposes of internal and external school accountability, depending on the way the models are specified. Following Raudenbush's (2004) conceptualisation and further extensions by Timmermans et al. (2011) on what school value-added models for accountability ought to include, this research explores different ways to inform school accountability to external stakeholders. Thus, a series of multilevel models are specified as an attempt to isolate the "true" school effects as reliably as possible by controlling for all the relevant non-malleable factors affecting pupils' progress in a realistically and sufficiently complex way, with the ultimate purpose of comparing schools fairly. This conceptualisation is then extended to inform internal school accountability by incorporating the analysis of a meaningful malleable factor affecting pupils' academic progress, such as cultural capital.

School accountability in Chile

Traditionally, government practice in Chile has employed school averages as indicators of effectiveness, although a new accountability system is being developed that takes into account some of the principles of value-added research (San Martín & Carrasco, 2012, 2013). This new system is to be implemented by the Agencia de Calidad de la Educación (Agency for the Quality of Education), from 2015 in a pilot stage. This methodology for classification of the schools will include a series of indicators of school quality, with attainment being the most important amongst them (Agencia de Calidad de la Educación, 2014). This quality assurance system will hold schools accountable based on the achievement of their pupils in Years 2, 4, 6, 8 and 10 in the SIMCE tests.

In this Quality Assurance System, schools will be classified into four categories: a) high performance; b) middle performance; c) lower-middle performance; and d) insufficient performance (Agencia de Calidad de la Educación, 2014). Two thirds of this classification will be dependent upon attainment scores averaged at the school level, which will be adjusted by a set of pupils' socio-economic and demographic characteristics to make allegedly fairer comparisons. Even though the specific details of the methodology have not been published yet, it has been announced that it will involve the implementation of multiple linear regression (MLR) models to make the necessary adjustments for the socio-economic and demographic characteristics of the pupils (Agencia de Calidad de la Educación, 2014). Although this is an improvement compared to current practice involving school averages only, the use of MLR models is indeed a major setback, given the overwhelming evidence in favour of the use of multilevel models. The advantages of using multilevel models over multiple linear regression models are discussed in more detail in chapter 3. Furthermore, the consequences of using biased models to assess school performance are analysed in more depth in chapters 4 and 5.

The issues around the particular methods to be applied are not trivial. This quality assurance system can be said to have adopted a high stakes testing approach (Au, 2007; San Martín & Carrasco, 2013), where schools persistently judged to be performing insufficiently could face closure. It is, therefore, key to ensure a fair system, where schools are held accountable for what they can actually act upon.

As mentioned before, even though school accountability is highly relevant for many different reasons pertaining diverse stakeholders external to the schools, its purpose is also arguably applicable to the schools themselves. Sustained school improvement and/or maintenance of standards require permanent internal monitoring and assessment of pupils' attainment and progress. A comprehensive account as such should consider the way in which external and internal factors affect pupils' academic success, in order to distinguish between factors on which the schools can and cannot intervene. The next section discusses how structural factors can shape pupils' educational performance and trajectories, with the ultimate purpose of defining a framework for the analysis of malleable factors susceptible of school intervention.

2.5. The socio-economic and cultural gradient in pupils' academic progress

The purpose of this section is to discuss how educational inequalities occur, searching for theoretical and empirical links between socio-economic structural factors and pupils' academic success. This is done with the ultimate purpose of informing internal school accountability and contributing to evidence-based school improvement intervention processes. This section begins by defining several general concepts related to how the education system is built, its purpose and function in society as a whole, its main characteristics, and its consequences on individuals; then, more specific concepts are discussed, such as cultural capital and its relationship with academic performance.

The institutionalised educational system started taking its current form during the early years of the nineteenth century. The process of industrialisation and the expansion of the cities in most Western societies are the main causes of the increased demand for institutionalised education and specialised training (Giddens, 1993).

Due to the foregoing, the main initial purpose of the education system was to create a mass of properly trained workers, who would have the capabilities and knowledge to sustain the massive production of the recently born industries (Durkheim, 1947; Giddens, 1993; Illich, 1973). Naturally, as societies develop, the specific objectives of schooling vary and thus, for instance, in the early stages of the development of the modern education systems, there was a strong emphasis on the acquisition of the abilities necessary for manufacturing (Durkheim, 1947; Giddens, 1993). As the necessities of the Industry grew, so did the diversification of the economies of industrialised countries, producing more differentiation in the education systems and far more specialisation (Durkheim, 1956; Giddens, 1993). Nowadays, there is a strong emphasis on the acquisition of 'soft' knowledge, that is the ability of learning to learn (especially at the highest levels of the system), because of the rise of the information society (See for example: Castells, 1996).

The development of education (or institutionalised education systems) has always had close links to the ideal of democracy. This is because of the opportunities that education provides to citizens to develop themselves, their abilities and aptitudes, which will eventually help reduce inequalities in Society regarding wealth and power (Giddens, 1993). This raises issues such as social selection and mobility as well as meritocracy, and so forth, which will be addressed in section 2.5.3.

There are a number of influential authors in the sociology of education literature, for instance, Bernstein (1975), who theorises about educational inequalities from a linguistic skills point of view. This author points out that pupils from diverse backgrounds acquire different 'language codes' that condition their academic path and performance. Another relevant and controversial author is Illich (1973), who introduced the notion of 'hidden curriculum', which refers to those non-stated or implicit objectives of schooling that have to do with discipline and hierarchy in society.

Another highly relevant and influential author in the Sociology of Education literature is Pierre Bourdieu (See for example: Bourdieu & Passeron, 1977; Bourdieu, 1988, 1990), whose most important contribution is an exhaustive description of the educational system from its foundations to its consequences to the reproduction of the social structures. The following sections focus mainly on the contributions of Bourdieuvian theory to the study of educational inequalities and school effectiveness.

2.5.1. The education system and the role of education in Society

As it has been mentioned above, the modern institutionalised education system was created with the purpose of preparing the young population for the requirements of the nations' economic needs. The underlying theoretical assumption here is that the education system fulfils a necessary function in the wider social system that allows the latter to perpetuate itself (Bourdieu & Passeron, 1977, 1979; Bowles & Gintis, 1976, 2002; Durkheim, 1947, 1956). In order to comprehend this function, it is necessary to define and analyse the main components of the education system.

Following Bourdieu and Passeron's seminal work (Bourdieu & Passeron, 1977, p. 6), the basic operation of the education system is the pedagogic action, defined as "(…) objectively, symbolic violence insofar as it is the imposition of a cultural arbitrary by an arbitrary power". In other words, the pedagogical action is the action by which an appropriate agent (a properly and institutionally trained and accredited individual) instructs or teaches certain socially approved and valued contents to an audience of learners (usually the youngest members of society). This action constitutes symbolic violence due to the fact that the power relations of any given society are the basis of the arbitrary power which underlies the establishment of the relations in the education system (Bourdieu & Passeron, 1977). In simple terms, the process of schooling is arbitrary (or symbolically violent), because the curriculum standards that a pupil needs to comply with are imposed conventionally as a result of the cultural hegemony of a particular socio-economic group (Bourdieu & Passeron, 1979; Bourdieu, 1988; Bowles & Gintis, 1976, 2002; Durkheim, 1956; Lareau & Weininger, 2003).

Furthermore, the pedagogical action constitutes symbolic violence considering that the imposed instruction of certain contents or concepts produces and reproduces the arbitrary selection of a certain social group or class (Bourdieu & Passeron, 1979). This is also the case for the examination of said contents, which must be perceived as "equitable" (Williams, 2012) to provide legitimacy to the achievement of the students and the education system in general. This selection of contents to be taught (curricula, in other words) is arbitrary in the sense that it does not derive nor can be deduced from any universal principle, set of values, etc.. Instead, it derives from the culture and set of valuable knowledge and skills defined by the dominant socio-economic group (Bourdieu & Passeron, 1979; Bourdieu, 1988).

With regard to its degree of arbitrariness, pedagogical action owes its power of imposition to the level of arbitrariness of the culture imposed (Bourdieu & Passeron, 1977; Bourdieu, 1977a). In other words, its arbitrariness depends on the power relations between the social groups, which are the reflection of the system of the cultural arbitraries of the dominant pedagogical action in any society or *habitus*, as termed by Bourdieu, i.e. the set of socially acquired dispositions or schemes of thinking and acting of the dominant social class (Bourdieu & Wacquant, 1992; Bourdieu, 1977b, 1990). This production and reproduction of the social structures, which are put in place by the mere functioning of the education system, are the essence of educational inequalities that are continuously recorded by educational research in school effectiveness studies, where the socio-economic background of pupils is found to influence their performance heavily. A discussion is presented in section 2.5.2 on how these socio-economic structures are present in educational studies and more specifically how these are taken into account in statistical models.

Furthermore, the establishment of an institutionalised education system is a recursive process, because its structural characteristics are the logical consequence of its own essential functions of instruction (and reproduction) of the cultural arbitrariness that allows it to sustain itself (Bourdieu & Passeron, 1979). This process is carried out in the most durable (and least expensive) manner, conforming as closely as possible to the dominant *habitus* by exercising an institutionalised pedagogical practice which is performed by a permanent corpus of specialised agents. These agents are homogenously trained and equipped with "standardised and standardising instruments" as termed by Bourdieu and Passeron (1979).

While establishing a standard pedagogic action, that is the process of production and reproduction of an arbitrary *habitus*, the institutionalised education system also sets the basis for questioning its own legitimacy. This derives in the necessity of creating the institutionalised conditions for disregarding the symbolic violence of the arbitrariness implied by its function (Bourdieu, 1988). Due to the fact that its mere functioning provides the sufficient conditions for legitimating its existence and persistence, the education system sustains itself through its apparent autonomy and neutrality, and is granted the monopoly of the legitimate use of symbolic violence. This symbolic violence inflicted by the education system needs to be perceived as "fair" and "equitable" to persist (Williams, 2012). This is known as the principle of dependence through independence (Bourdieu & Passeron, 1977; Bourdieu & Wacquant, 1992; Bourdieu, 1990). In plain terms, the system depends on its independence.

As mentioned previously, the development of the education system has been tightly related to the development of the economies of the industrialised countries. This raises issues around the extent to which the ideals of personal development, democracy, building citizenship have been achieved (Byrne & Rogers, 2000). Following that line of thought, the modern education system should be understood not only to be merely related to the needs of the economy, but also as a response to them (Illich, 1973).

According to Illich (1973), schools have developed to fulfil four basic social functions, namely: the provision of custodial care for children that derived from the imposition of mandatory education; the selection, preparation and distribution of people amongst occupational roles; the learning of dominant values, norms and culture; and the acquisition of the socially approved and most valued (from a dominant culture's point of view) skills and knowledge.

This vision is consistent with Bowles and Gintis (1976) who assert that schools "are destined to legitimate inequality, limit personal development to forms compatible with submission to arbitrary authority, and aid in the process whereby youth are resigned to their fate" (p. 266). Later, when revisiting their previous research (Bowles & Gintis, 2002), these authors also stress the relevance of the inter-generational transmission of socio-economic status in the process of schooling, and the relationship between the latter and labour market success. In the next section, it is appreciated that parents' income and educational level are two highly relevant variables that are frequently included in statistical model for assessing educational outcomes in school effectiveness studies. The next section discusses the role played by socio-economic structural factors in the process of schooling, focusing on how these external factors affect pupils' academic attainment and progress.

2.5.2. The influence of the socio-economic background in educational trajectories

As discussed in the previous section, the schooling process cannot be comprehended separately from the structural processes that a society is permanently producing and reproducing. In this respect, the notion of cultural capital is crucial. From this viewpoint, school success is the effect of socio-economic advantages in the education system, as well as in society itself.

Cultural capital is a highly abstract concept that might be understood as a set of "instruments for the appropriation of symbolic wealth worthy of being sought and possessed" (Bourdieu, 1977a, p. 488). These instruments, in generic terms, refer to any given competence, knowledge, skill, ability, way of thinking or acting, etc., that enables the appropriation of the cultural heritage of a social formation or group by any given subject (Lareau & Weininger, 2003, 2004; Robbins, 2005). This cultural capital might take three forms in the social world, namely: the embodied form of long-lasting dispositions of the individual, the objectified form of cultural goods and the institutionalised form of academic qualifications (Bourdieu, 1986).

Following this conceptualisation, the magnitude and value of the cultural capital acquired by a particular individual depend on the social contexts to which this individual has to cope with. Therefore, primary socialisation is highly relevant because of its fundamental role in providing the youngest members of society with the capital they need to perform in school, as well as in the labour market, and, more generically, the social world (Bourdieu & Wacquant, 1992; Bourdieu, 1990).

The distribution of the cultural capital in any given society is unequal and unfair (Bourdieu & Passeron, 1979; Bourdieu, 1988). The accumulation of cultural capital depends on the socio-economic status of its members, as well as the value of the cultural capital inherited during primary socialization. This, in turn, depends on the distance of the relative social position of the family background of a pupil to the relative position of the predominant social class or group (Bourdieu & Passeron, 1977, 1979; Bourdieu, 1977a). This causes the educational system to create differentiated levels of demands for pupils according to their socio-economic background (Bourdieu & Passeron, 1979; Bourdieu, 1988). This is because the standards in an educational system are arbitrarily imposed in the same fashion as curricula, and they are highly similar for every pupil. In consequence, given that pupils have different levels of appropriation of the most socially valued knowledge and skills before the initiation of their schooling process, and that *in itinere* there is an unequal access to educationally relevant resources, the demands from the school system to disadvantaged pupils is relatively much higher compared to the demands for the non-disadvantaged pupils (Bourdieu & Passeron, 1977, 1979; Bourdieu, 1988).

Consequently, pupils belonging to the most economically challenged social groups, with the lowest levels of the highest valued cultural capital in a society, are those who are compelled to go through the most demanding and rigorous processes of educational selection throughout the whole system (Bourdieu & Passeron, 1979). Moreover, this implies that these "disadvantaged" pupils need to undergo a process of "acculturation" (Bourdieu & Passeron, 1979), that is the acquisition of a large amount of strange cultural capital, which makes them more likely to fail than to succeed in the school system. In contrast, economically, and hence, culturally advantaged children are much more likely to succeed than to fail in this selection process (Bourdieu & Passeron, 1977, 1979). This topic will be further developed in the next section.

These fairly well-known educational inequalities have been a great concern for practitioners, policy-makers and researchers during the last few decades. Pupils' academic attainment and progress has been studied empirically using the concept of cultural capital as the link between the socio-economic background and academic success. Many studies (See for example: De Graaf, 1986; Di Maggio & Mohr, 1985; Di Maggio, 1982; Dumais, 2002) have attempted to analyse the relationship between cultural capital and educational attainment by means of various operationalisations of cultural capital focused mainly on its "objectified form". Authors such as Lareau and Weininger (2003) and Kingston (2001) have heavily criticised these studies, claiming that Bourdieu's concept of cultural capital has been unnecessarily and superfluously reduced to social status attainment by means of the acquisition of prestigious cultural goods or resources, and participation or interest in "highbrow" culture.

On another front, the concept of cultural capital itself has been heavily criticised by many authors. For instance, Edgerton et al. (2012) and Edgerton and Roberts (2014) argue that the ample variety of ways in which cultural capital (and other Bourdieuvian concepts) has been operationalised has arisen because of Bourdieu's imprecise definitions. Nonetheless, these authors provide an auspicious view on the use of Bourdieu's concepts in educational effectiveness research, advocating for the analysis of cultural capital, along with other Bourdieuvian concepts such as habitus, field and practice, as powerful theoretical and empirical tools. These authors argue that, despite fair criticism, Bourdieu and Bourdieu-inspired subsequent theorists have set up a conceptual framework for the study of persistent educational inequalities.

Moreover, Sullivan (2001, 2002) stresses the vagueness of Bourdieu's definition of cultural capital. This vagueness has induced heterogeneous operationalisations in empirical research to such degree that evidence is not always consistent across studies. Nevertheless, this author acknowledges the insight that the concept can provide and its potential usefulness for explaining, at least partially, the influence of socio-economic structural factors on pupils' academic success. The author argues that some operationalisations of cultural capital have successfully proved to be useful for analysing pupils' educational outcomes; cultural participation, for instance, operationalised in parents' and pupils' reading behaviour, as well as the number of books at home, were found to be positively associated with better outcomes. In chapter 6, the variables "number of books in the household", as well as pupils' and parents' reading frequency are considered as realisations of the latent variable cultural capital, which will be estimated to explore models that can be useful for schools to modify their policies towards improvement.

Furthermore, Sullivan (2002) argues that the effect of cultural capital on pupils' performance may also vary across countries at different times, i.e. cultural capital may be more important in some countries than others or it may operate in different ways according to context. This last argument is especially relevant for this research, because it may be claimed that the context of developing countries may moderate (either upwards or downwards) the effect of cultural capital on pupils' academic success. In simpler terms, the effect of cultural capital might be completely different from what is found in developed countries. This is why it may be worthwhile exploring this when searching for better ways to inform internal school accountability for school improvement purposes.

2.5.3. The relationship between socio-economic background and selection in the education system

As a consequence of its role in society, the institutionalised educational system plays a key role in fostering socio-economic class membership through the provision and certification of cultural competences, which are arbitrarily assumed to be of greater importance or value (Bourdieu & Passeron, 1977, 1979; Bourdieu, 1977a). Given the latter, the familiarity with the culture of the dominant social class, i.e. the result of the pedagogical action carried out by the primary social context (family of origin), provides highly important advantages to students in the education system (Lareau & Weininger, 2003, 2004; Williams, 2012). In other words, the social, cultural and economic context in which the primary socialisation takes place, strongly conditions the academic performance of pupils.

> "Given that they have had to achieve a successful acculturation in order to meet the irreducible minimum of academic requirements [...] the working-class and middle-class students who reach higher education have necessarily undergone more stringent selection […]" (Bourdieu & Passeron, 1977, p. 73).

The arbitrariness (seeming autonomy and neutrality) of the selection of contents to be taught and the standards to achieve in the school system has substantial consequences on pupils. One of them is that pupils who proceed from low socio-economic contexts often justify their low achievement by describing themselves as 'not clever enough', tending to disregard their 'street knowledge' (Willis, 1977). This is because of its alleged worthlessness in the economic system.

Likewise, the role of occupational aspirations is highly relevant to academic success (Halpern, 2005; Hernandez-Martinez et al., 2008; Williams et al., 2009). Often children from low socio-economic backgrounds tend to adjust, justify or resign themselves to their poor academic performance according to a set of occupational aspirations. These aspirations are socio-culturally constructed insofar as they are in line with what their parents and the wider social environment (including schools) have contributed to mould during the process of schooling (Hernandez-Martinez et al., 2008; Williams et al., 2009).

Furthermore, the effect of the parents' educational level as a proxy for cultural capital has been explored extensively. Several authors (see for example: Feinstein et al., 2004; Lareau & Weininger, 2003; Murillo & Román, 2011) have reported a positive association between parental qualifications and occupation and their children's academic performance. The underlying cause for this relationship is that parents with a higher educational level are arguably better equipped to attend to the pupils' needs, assist and guide them, get involved in the school environment and instil them to achieve higher, and eventually pursue higher education (Edgerton et al., 2012; Edgerton & Roberts, 2014; Kleanthous & Williams, 2013; Lareau & Weininger, 2003). This argument is key in this research; in chapter 6, the variable parents' qualifications is added as one of the manifest variables to estimate the measure of cultural capital, which is used to create a model for informing internal school accountability.

To sum up, students who proceed from the most advantageous socio-economic contexts have a 'head start' simply by being immersed since birth into the predominant culture. Furthermore, they have even greater advantages when one takes into account the purchasing power and access to meaningful resources to succeed in the education system. In contrast, those students proceeding from low socio-economic backgrounds are obliged to undertake a more demanding educational path, due to the fact that they need to interiorise a symbolic system (contents, curricula, standards to achieve) that they are unfamiliar with. Regretfully, they have to undergo this process with a much more restricted access to educationally-relevant resources than their more socio-economically advantaged counterparts.

2.6. The Chilean education system

The Chilean education system is regulated by the General Act of Education №20370 of 2009[4], which is the product of a series of reforms, starting in the 1980's during Pinochet's Dictatorship (1973-1990). The current form of the education system is to be maintained until 2017, when significant reforms are to be implemented, including the modification of the structure of grades for primary and secondary education and the regulation of selectivity, as well as the reform to the shared funding scheme in publicly funded schools. Currently, education in Chile is compulsory from the age of six until the age of 18. The total of 12 years of compulsory education is divided into primary and secondary education. Primary school is comprised of eight grades, which are classified in six 'Basic Levels' as seen in Table 2.1.

Table 2.1: Structure of the primary education in Chile

Basic Level 1	Years 1 and 2
Basic Level 2	Years 3 and 4
Basic Level 3	Year 5
Basic Level 4	Year 6
Basic Level 5	Year 7
Basic Level 6	Year 8

On the other hand, secondary education is comprised of four grades and is subdivided in two types of curricula, namely, 'Scientific-Humanist' secondary education and 'Technical-Professional' secondary education. In the 'scientific-humanist' path, pupils are taught a curriculum with subjects that prepare them in a wide range of general skills and knowledge, which are considered as the basics for Higher Education. Pupils usually make their choice among a 'science path' to undertake studies in subjects such as Mathematics, Physics, Chemistry and Biology; or a 'humanities path' to undertake studies in Literature, History and Philosophy. With respect to the 'Technical-Professional' path, pupils undergo a similar curriculum to that of 'Scientific-Humanist' education during the first two years, and in the last two years, they receive vocational training of their choice from a wide range of working skills,

[4] General Act of Education (Ley General de Educación), Act №20370, 2009 (Chile). This Act rescinded the predecessor Act №18962 of 1990. http://www.bcn.cl

such as mechanics, electricity, accountancy, gastronomy, nursery, and so forth. This can be appreciated in Table 2.2.

Table 2.2: Structure of the secondary education in Chile

Paths	First cycle (Years 9 and 10)	Second cycle (Years 11 and 12)
Scientific-Humanist	Common curriculum in generic skills and knowledge	Choice of subjects in Science or Humanities
Technical-Professional		Choice of Vocational training

Regarding the administrative aspects, all schools (secondary and primary) in the Chilean education system can be divided into three major institutional types, which are described below.

a) State-funded schools

These schools are fully funded by the State through a system of subsidies or vouchers, which are paid by the State as if they were tuition fees. These subsidies are fixed for every pupil and are subject to pupils' attendance to the school, which implies that the funds received by schools under this administrative regime vary. Preferential subsidies have been implemented in the last decade (150% of the regular subsidy per pupil) to those pupils who are in a more vulnerable socio-economic situation, or are in risk of becoming school-leavers. These institutions are usually managed by the City Council, or are delegated to a Municipal or Local Board of Governors. These schools follow the National Curriculum in full; they are comprehensive and fully supervised by the Ministry of Education.

b) Subsidised independent schools

These schools are partly funded by the State and private parties, and in some cases, parents contribute to the funding by the payment of fees. The funds delivered by the State to these institutions follow the same scheme of state-funded schools, i.e. there are fixed vouchers per pupil and preferential vouchers for vulnerable pupils. This type of schools may also be divided into two categories: a) fee-paying, and b) non-fee-paying. Regarding their administration, these schools may be managed by private parties, such as religious institutions, Corporations, Societies, entrepreneurs, etc. These schools need to follow the National Curriculum, which is usually considered as the minimum standard. They are usually selective (although there are no precise data on this matter) and they are supervised by the Ministry of Education.

c) Independent schools

These schools are funded by private parties and the contribution of parents through fees and donations. Fully independent schools are administered by private parties, do not receive funds from the State, and are only subject to partial supervision from the Ministry of Education. This category also includes those schools with a delegated administrative regime, which corresponds to schools managed by private parties and partially by the State via public contracts. Furthermore, Independent schools are not obliged to follow the National Curriculum and are highly selective.

Table 2.3: Pupils enrolled in Chilean primary schools from 2004 to 2013, sorted by school type.

School type	State-funded		Subsidised		Independent		Total
Year	N	%	N	%	N	%	N
2004	1,181,833	52.39%	911,758	40.41%	162,399	7.20%	2,255,990
2005	1,139,100	51.47%	933,468	42.18%	140,642	6.35%	2,213,210
2006	1,081,699	49.80%	952,265	43.84%	137,950	6.35%	2,171,914
2007	1,028,517	48.28%	961,789	45.15%	139,937	6.57%	2,130,243
2008	976,968	46.60%	978,266	46.66%	141,238	6.74%	2,096,472
2009	935,045	45.05%	998,014	48.08%	142,523	6.87%	2,075,582
2010	881,658	43.42%	1,005,486	49.52%	143,214	7.05%	2,030,358
2011	843,378	42.40%	1,002,612	50.40%	143,165	7.20%	1,989,155
2012	810,170	41.29%	1,010,239	51.48%	141,845	7.23%	1,962,254
2013	785,042	40.37%	1,013,514	52.12%	146,083	7.51%	1,944,639

Source: Chilean Ministry of Education, Centro de Estudios MINEDUC (2014)

In Table 2.3, it is appreciated that the ample majority of Chilean pupils in primary education attend either state-funded schools or subsidised independent schools. More interestingly, in this period of 10 years, it is appreciated that State-funded schools have been steadily decreasing in the number of student they serve; meanwhile subsidised independent schools have been steadily increasing. In the whole period, a shift is observed from the majority of pupils attending State-funded primary schools in 2004 (52.39%), towards the majority of pupils attending subsidised independent schools in 2013 (52.12%). On the other hand, the relative participation of independent schools in the system has remained stable throughout the 10-year period. In Table 2.4, a similar trend is observed regarding secondary school education in Chile.

Table 2.4: Pupils enrolled in Chilean secondary schools from 2004 to 2013, sorted by school type.

School type	State-funded		Subsidised		Independent		Total
Year	N	%	N	%	N	%	N
2004	450,477	45.67%	404,369	41.00%	131,456	13.33%	986,302
2005	461,706	45.01%	432,782	42.19%	131,222	12.79%	1,025,710
2006	453,352	43.62%	456,269	43.90%	129,816	12.49%	1,039,437
2007	440,051	42.73%	460,153	44.68%	129,686	12.59%	1,029,890
2008	422,070	41.56%	464,516	45.74%	128,872	12.69%	1,015,458
2009	410,916	40.85%	467,704	46.50%	127,195	12.65%	1,005,815
2010	392,421	39.57%	471,979	47.59%	127,316	12.84%	991,716
2011	376,632	38.43%	476,189	48.59%	127,166	12.98%	979,987
2012	345,630	36.81%	471,255	50.19%	122,051	13.00%	938,936
2013	335,769	36.30%	467,375	50.53%	121,761	13.16%	924,905

Source: Chilean Ministry of Education, Centro de Estudios MINEDUC (2014)

Table 2.4 displays the relative participation of the three school types in the secondary education system throughout the period 2004-2013. As is the case of primary school education, it is noted that in secondary school education, during the last 10 years, the total number of pupils in State-funded schools has been steadily decreasing. Meanwhile, the number of students attending subsidised independent schools has been steadily growing from a 41% in 2004 to just above half (50.53%) the pupils in 2013. As noted also in Table 2.3, the total number of students attending independent secondary schools has remained stable in this 10-year period.

Paredes and Pinto (2009) argue that this downward tendency for the number of pupils enrolled in State-funded schools may be reflecting upon the Chilean economic growth. They argue that higher income conveyed an increase in the demand for subsidised independent schools with a shared funding scheme, where parents could contribute economically towards a better quality of education. A less restrictive administrative regime might have also helped subsidised independent schools thrive in the last decade.

The socio-economic gradient in the Chilean education system has been analysed more extensively by educational researchers in Chile from the beginning of the 2000's. Until 1997, the Chilean Ministry of Education had no records of attainment at the pupil level and hence early studies of educational effectiveness used only aggregated school data (Mizala & Torche, 2012). A further obstacle, as pointed out earlier (section 2.3.5), is that pupils only sat the SIMCE tests once throughout their whole educational trajectory until 2006. Until then, pupils used to sit the SIMCE tests either in Year 4, Year 8 or Year 10. Nonetheless, Chilean-based research has, one way or another, found evidence of deep educational inequalities due to the socio-economic background.

Valenzuela et al. (2008) investigated the segregation of the Chilean school system more deeply using the dissimilarity index (Duncan & Duncan, 1955). The degree of dissimilarity was measured by comparing the proportion of disadvantaged pupils with the proportion of non-disadvantaged pupils within schools. Disadvantage or vulnerability is measured by household income and parents' educational level. The authors found that State-funded schools are far less segregated than subsidised independent schools, which in turn are also far less segregated than independent schools; moreover, subsidised independent schools with a shared funding scheme are more segregated than non-fee paying subsidised independent schools. Furthermore, the authors found that the socio-economic segregation of the schools significantly varies across areas, depending on the proportion of people living in poverty. The authors concluded that the segregation of the Chilean education system as a whole is extremely high and this has been steadily increasing in the period between the years 1998-2006, mainly because of the existence and growth of subsidised independent schools and especially those with a shared funding scheme, which are the ones that can more easily shape pupil intake via the imposition of fees.

Given the high level of socio-economic segregation of the school system, differences in pupils' achievement according to school type are expected to be large. Using school aggregated data, McEwan and Carnoy (2000) found that subsidised independent schools did not perform better than State-funded schools on average when resources were at similar levels. McEwan (2001) further investigated the differences between State-funded schools, subsidised independent schools with Catholic religious denomination, subsidised independent school with no religious denomination, and independent schools. The author found evidence that independent schools, and religious schools to a lesser extent, outperformed State-funded schools; while non-religious subsidised independent schools were found to be somewhat less effective. The author controlled for explicit selectivity by specifying controls for pupils' socio-economic factors and prior attainment and unobservable selectivity, using Heckman's (1979) bias correction. Thus, it was found that, when controlling for selectivity, the effect sizes are smaller for all schools and the differences are less pronounced. Tokman (2002) takes a step further in arguing that, after controlling for non-random selectivity, State-funded schools are not uniformly less or more effective than subsidised independent schools, but rather they are more effective for pupils from deprived socio-economic backgrounds.

Mizala and Romaguera (2000, 2001) provide further evidence of the large differences in the relative effectiveness of schools according to their institutional type. They also point out a sizable relationship between socio-economic status and school type, which denotes a noticeable segregation of the school system. These authors also highlight that subsidised independent schools are, in terms of effectiveness, closer to State-funded schools than to independent schools, because analyses show that, after controlling for socio-economic background, these two school types do not significantly differ. Furthermore, the gap between independent schools and the other school types is still significant after controlling for socio-economic background; however, it is greatly reduced.

The interest in studying the differences in the educational outcomes achieved by the diverse school types has its origins in the profound impact of the educational reforms carried out during the Pinochet's Dictatorship (1973-1990) from 1981 onwards. The foundations of the current education system were built during this period and the creation of subsidised independent schools was the most relevant reform. Torche (2005) analysed the impact of the reforms on the educational transitions from primary school all the way through tertiary education, in seven birth cohorts from 1936 to 1976. Comparisons of the outcomes of the younger cohorts and the older cohorts indicate that the reforms had a noteworthy impact on educational transitions. Evidence of growing inequalities was found, where individuals from the younger birth cohorts who attended independent schools had distanced themselves (had greater chances of achieving further levels of education) even further from individuals who had attended State-funded schools than individuals from the older birth cohorts who attended independent schools. It was also found that individuals who had attended subsidised independent schools had greater chances of graduating from secondary school than their peers in State-funded schools. In sum, the emergence of subsidised independent schools during the 1980's reforms had a catalysing effect on educational inequalities.

More recently, Mizala and Torche (2012) analysed the socio-economic distribution of achievement within and between schools in Chile using data from the SIMCE database (2002-2004). Examining pupils' attainment in Mathematics and Spanish Language, the authors found that subsidised independent schools have a larger between-school variance than State-funded schools, suggesting that while they serve a more heterogeneous student body, subsidised independent schools in particular serve specific homogenous socio-economic groups. The authors claimed that flexible regulations have allowed subsidised independent schools to shape their student bodies and teaching staff to specialise in particular market niches. They particularly highlighted the shared funding scheme (fees paid by parents and carers) in subsidised independent schools, as the main driver of school socio-economic sorting. The impact of this school sorting phenomenon was found to be more severe in subsidised independent schools than in State-funded schools. Even though pupils in either State-funded schools or subsidised independent schools were substantially affected by socio-economic status, either that of their families' or their schools', it is the pupils in the latter type of schools the ones whose attainment was more strongly affected by the socio-economic context of the school.

In conclusion, socio-economic segregation in the Chilean education system is closely associated with the characteristics of its institutional design. The mixture of partially and fully publicly-funded as well as fully privately-funded schools has deeply contributed to the persistence (and even growth) of inequality in educational outcomes. Although, individual (or more precisely, family) level socio-economic status continues to be a relevant source of variation in pupils' academic achievement, the effect of the school socio-economic context is even greater. Its effects are by all means considered in this research and examined thoroughly in the construction of the contextualised value-added models for internal and external school accountability. In the next section, a summary of the key concepts discussed in this review is presented.

2.7. Summary of key concepts

Throughout this literature review, a number of key concepts have been described and some of their interrelationships have been discussed. The purpose now is to summarise these topics and concepts as they will be used for the remainder of this thesis. Table 2.5 presents a summary of the key concepts.

Table 2.5: Summary of key concept used throughout this thesis.

Concept	Description
School value-added	This is an indication of the extent to which a particular school has contributed to the progress of its pupils, beyond what is expected of them, considering their initial achievement and their socio-economic background.
School effectiveness	This concept can be used interchangeably with the concept of school value-added, as it indicates how effective a particular school has been in fostering the progress of its pupils.
Pupils' attainment	This is a raw measure of the level of pupils' achievement at a particular academic subject with respect to certain objective criteria, which are usually standard goals set in the National Curriculum.
Pupils' progress	In its simplest form, pupils' progress can be defined as the difference between prior and subsequent attainment. However, from a value-added perspective, this is incomplete. In this thesis, progress is a measure that compares pupils' prior attainment in a particular subject with pupils' subsequent attainment in the same subject, also considering the pupils' socio-economic background. Hence, pupils' progress would constitute the difference between what is expected of a pupil and what a pupil achieves, conditional upon their prior attainment and their socio-economic background.
CVA model	A contextualised value-added model is a statistical model that attempts to isolate the "true" schools' contribution to the progress of their pupils, considering their prior attainment and socio-economic background. This statistical model is most often a multilevel model, where the total variance in a particular subject is split between (at least) pupils and schools, and their socio-economic characteristics are used as explanatory variables to "contextualise" the information.
School accountability	This is the system by which schools are held responsible for their performance, i.e. the progress of their pupils. Depending on the recipients of the information, school accountability can be understood as: a) external, when the information is to be used by external stakeholders, such as government officials, for policy-making purposes; and b) internal, when the information is to be used by school authorities for improving their own practices.
School accountability measures	These are the indicators used to judge the performance of schools, i.e. the contribution they have made to their pupils' progress. These measures are also known as CVA scores or CVA estimates. These indicators correspond to the residual point estimates at the level of schools that are derived from the multilevel models used to analyse pupils' progress. More details of this are provided in chapter 3.
Cultural capital	This abstract concept refers to the set of knowledge, skills, abilities, etc. that allows the bearer to seek and possess the most valued cultural heritage of a Society, i.e. the cultural assets and knowledge that are most prevalent in the highest socio-economic classes. Cultural capital is theorised to take three basic forms in the social world: a) the objectified form of cultural goods; b) the institutionalised form of academic certifications; and c) the embodied form of long-lasting dispositions.

In the final section of this chapter, the research questions are unfolded and briefly discussed in light of the literature review.

2.8. Research Questions

This research will seek to answer the following questions:

1. What are the effects of classrooms, local authorities and primary schools on Mathematics and Spanish Language progress in Chile, beyond pupils' socio-economic and demographic characteristics and secondary school-level effects?

This first question seeks to ascertain whether there exist additional sources of variation that go beyond what has been traditionally analysed in school effectiveness research. As mentioned earlier, school value-added models have traditionally assessed the variation in standardised test scores arising from only two sources: the schools and the pupils. However, there is evidence from previous research that statistical assumptions underlying such models are untenable and they obscure substantially relevant relationships. This research question is addressed more deeply in chapter 4.

2. How are pupils' progress and school value-added in Mathematics related to pupils' progress and school value-added in Spanish Language in Chile?

Further to the first research question, school effectiveness has also been traditionally analysed in diverse subjects that have been treated in a dissociated way. As mentioned previously in this chapter, this is in spite of evidence supporting the idea of school effectiveness being a multidimensional phenomenon, insofar as neither pupils learn subjects completely separated from each other, nor do schools teach subjects in such a way. Neither in the learning nor in the teaching process do the connections between subjects need to be explicit or intentional. This is equivalent to assert that pupils who perform well in one subject can be expected (within a reasonable margin of error) to perform well in other subjects, after controlling for other relevant factors. Meanwhile, schools are reasonably expected to teach all subjects within the same standards; this implies that schools doing well in one subject would also be expected to be doing well in other subjects. Chapter 5 focuses on addressing this research question.

3. What relevant (non-negligible) differences can be found between school accountability measures derived from diverse statistical models?

The third research question derives logically from the previous two questions. Given that traditional models are allegedly inferior and that further adjustments are necessary for a more reliable estimation of school effects, it is worth investigating the usefulness of such extended (and more complex) models. School effectiveness models are recurrently used to inform the public about school performance for the main purposes of accountability and school choice via the construction of performance rankings. This research question aims to ascertain whether substantively relevant information can be extracted from the extension of the traditional models and whether this information is sufficiently different from what can be derived from simpler models. This research question is tackled throughout chapters 4 and 5.

4. How does cultural capital along with socio-economic and demographic characteristics of the pupils and the schools affect pupils' progress and school effectiveness in Chile?

This final research question is devoted to delve into the analysis of pupils' heterogeneity in Spanish Language and Mathematics attainment for the purpose of internal school accountability. In previous analyses, the focus was set on the between-school variability, controlling for all relevant external factors, including schools' context, geographical location, within-school variability, previously attended schools, etc, with the purpose of effectively isolating school value-added to inform external school accountability. After having controlled for those factors, the major source of variation in academic progress lies in the pupils' characteristics and abilities which previous models have not accounted for. Some of those unaccounted pupils' characteristics are deemed to be malleable while others are not. The theoretically relevant element of this analysis is that cultural capital is a malleable factor that could boost children's performance and whose influence can be controlled at various levels, i.e. via school policy, local authority initiatives, central government public policy-making, private initiatives, as well as parental courses of action. This is the main focus of chapter 6 of this thesis.

In the next chapter, the data used in this thesis are described and the methods to analyse them are outlined.

Chapter 3: Data and Methods for analysing school-value added

3.1. Introduction

The main aim of this chapter is to describe the methods by which the research questions will be addressed and the data that will be analysed to address these questions in subsequent chapters. This chapter starts off with describing the data used in this research, focusing on the main variables that were selected to implement the statistical models of chapters 4, 5 and 6. Afterwards, the methods will be presented, beginning with the most basic statistical models that can be used to analyse pupils' performance in standardised tests, i.e. multiple linear regression, and discuss its pitfalls, which serves as an introduction to discuss the statistical superiority of multilevel modelling. Then, the chapter moves on towards describing the multilevel modelling framework, its basic principles and characteristics, as well as how different specifications can help explore diverse research questions, and more specifically, the way in which the models fitted in this research are built. The equations corresponding to each diverse specification are presented and discussed. It will be shown how the most basic multilevel models can be extended to cater for the needs of analysing intrinsically complex phenomena such as school value-added and pupils' progress.

By the end of this chapter, the analytical strategy unfolds the way in which the selected variables are to be used to build the models; describing the steps to be taken in detail. Finally, some ethical considerations are briefly discussed.

3.2. Data

The data have been provided by the Chilean Ministry of Education. The SIMCE tests assess the achievement of core areas of the National Curriculum and are taken every year by Year 4 pupils and Year 8 pupils or Year 10 pupils.

This thesis draws on analysis of the cohort of pupils who sat the tests in 2006 when they were in Year 10 or 2nd year of Secondary School and also sat the tests previously in 2004. About 84% of the pupils who sat the test in 2006 were traced back to 2004. The expected matched cases rate is approximately 85%, because the drop-out rates are 5% and the retention rates are 10% (Centro de Estudios MINEDUC, 2014).

3.2.1. The SIMCE database

SIMCE[5] is a series of standardised tests created during the decade of the 1980's with the purpose of measuring the level of pupils' achievement of the goals defined in the National Curriculum. The tests evaluate the level of achievement of the Fundamental Objectives and the Minimum Compulsory Contents[6] defined for every subject in the National Curriculum. The current format of the SIMCE tests was consolidated in the year 1998[7].

The main subjects evaluated in these tests are 'Mathematics' and 'Spanish Language and Communication', which have been measured throughout the entire history of the testing system. Other subjects evaluated in the SIMCE tests have included: 'Social and Natural Sciences', 'English Language', 'Physical Education' and 'Information and Communication Technologies Proficiency'[8].

SIMCE test are taken every year, since 2006, by all students in the 4th grade and all students either in the 8th grade or in the 10th grade. Given this staggered pattern in the application of the tests, only recently, pupils started to sit the test more than once. The particular cohort to be analysed in this research is comprised of those pupils who sat the SIMCE tests in 2006 when they were in the 10th grade (or 2nd year of secondary school) and had previously sat the SIMCE tests in 2004, when they were in the 8th grade (last year of primary school). This is the only cohort moving from the last year of primary school to secondary school for which data are currently available. Using data from this cohort is particularly relevant for this research, since these pupils have fully undergone primary school, and hence its carry-over effects towards attainment in secondary school can be estimated.

3.2.2. Construction of the data sets used for the analyses

The cohort subject to analysis in this research is comprised of all Year 10 pupils who sat the SIMCE test in 2006. Most pupils from this cohort have also sat the SIMCE test in 2004, when they were in Year 8. In table 3.1, the case processing summary is reported for both occasions. It is observed that the ample majority of cases were uniquely identified with their anonymous identity number. The reasons for duplication or missing identifiers that the Chilean Ministry of Education have reported are related to illegible handwriting, mistakenly recorded identification numbers and blanks.

---

[5] SIMCE stands for "Sistema de Medición de la Calidad de la Educación". This can be translated as: Measurement System for the Quality of Education.
[6] Source: http://www.simce.cl
[7] Ibid.
[8] Ibid.

Table 3.1: Case processing summary for Cohort 2004-2006

Occasion year	2004	2006
Total number of cases	279,437	243,337
Number of valid cases†	277,679	242,306
Percentage of valid cases‡	99.22%	99.06%

† A valid case is one whose anonymous identifier is not duplicated or missing within a particular occasion.
‡ With respect to the total number of cases within a particular occasion.

It is also appreciated that the total number of cases, as well as the total number of valid cases, are lower in the second occasion of the test. There are a number of reasons for these differences. Year repetition of pupils is one of the most frequent: pupils who sat the tests in 2004 and were held back during 2004 and 2005 would not sit the SIMCE tests in 2006; this number is unknown since there are no records for them in 2006. Conversely, some pupils who sat the tests in 2006 and were held back before, were prevented from sitting the SIMCE tests in 2004, depending on the exact year in which they were held back and the number of times this occurred. This number is unknown since there are no records for them in 2004. Other reasons for the difference in the total number of cases in both occasions have to do with school dropout; there is expected to be an important dropout rate when pupils move from primary to secondary school. Furthermore, other pupils simply might have not attended school on the day of the tests. This could also be the case of schools attempting to "game the system" by making low-attaining pupils refrain from attending to gain undue advantage. This is a very plausible occurrence; however, it is undocumented, and hence, beyond the control of the statistical models.

According to the Chilean Ministry of Education (Ministry of Education - Chile, 2015), in 2010, the overall year repetition rate in primary schools was 4.3%, while in secondary schools, the rate was 8.1%. Likewise, the average dropout rate in primary schools was 5.5% whereas in secondary schools the rate was 10.7%. These rates are useful to understand attrition in the SIMCE tests; however, there are no precise records of the reasons for not sitting the SIMCE tests in the data sets. Unit non-response could eventually be a source of bias in the results; however, there are grounds to consider this issue as not overly problematic for the purpose of this thesis, because expected dropout (from school) and year repetition rates are consistent with the number of unmatched cases. The analysis of the causes of school dropout and/or year repetition are beyond the scope of this thesis. Issues around missing data are discussed in more detail in later sections 3.7.3, 8.5, and appendix 4.

In total, 202,605 pupils (approximately 84% of the total number of valid cases in 2006; see: Table 3.1) who sat the SIMCE tests in 2006 can be traced back to 2004. The percentage of matched cases has been calculated with respect to the overall number of valid cases in 2006, because this is considered to be more informative to the statistical model that intends to predict the scores in the second occasion of the tests.

3.2.3. Variables used to implement the CVA models

The variables to be used in the implementation of these CVA models can be classified in four categories, namely: a) attainment (outcome) variables, which are presented in Table 3.1; b) explanatory variables at the pupil and school level, which are presented in Table 3.2; c) variables associated to cultural capital, which are presented in Table 3.3; and finally d) level identifiers, which are presented in Table 3.4.

The choice of explanatory variables at the pupil and school level is rooted in the literature, following the principle of controlling for relevant variables which schools cannot modify freely in order to isolate as much as possible the "true" school effects. Prior achievement in Mathematics and/or Language is a prerequisite for fitting any value-added model insofar as it allows estimating the progress that pupils have made in those two subjects. With respect to the pupil-level variables, gender and family income escape the schools' control and they have been found to impact on pupils' performance largely. Pupils' year retention is a variable over which schools have partial control, since some schools can choose not to admit such pupils, and therefore, it may be a useful way to control, at least partially, for implicit school selectivity practices in the same way that prior achievement does (San Martín & Carrasco, 2013). This is debatable and other variables should be included, such as the number of students who have applied to a certain school and have been rejected or the number of students who have sat selectivity tests; however, there are no reliable measures of explicit school selectivity practices in the SIMCE database.

With respect to school-level variables, the socio-economic composition is a relevant influential factor identified in previous research; therefore, School SES was included in the models. In addition, school type, as Timmermans et al. (2011) point out, is a non-malleable school characteristic that may be considered in type X value-added models. This is also a debatable choice; however, the stance assumed in this research is that schools cannot freely switch into a different institutional type, and hence, a fair comparison should control for this, even at the expense of partially obscuring effects truly proceeding from diverse school practices associated with this.

As described in chapter 2 (section 2.6), Chilean schools are classified in 3 main types: a) State-funded schools; b) subsidised independent schools; and c) independent schools. Over the years, the coexistence of these three school types has produced a high level of segregation on the Chilean education system (Mizala & Torche, 2012; Torche, 2005), which is why it is relevant to ascertain the diverse effects that these schools have on pupils' performance. More details of the variables used in this thesis are given in Tables 3.2, 3.3, 3.4, 3.5 and 3.6.

Table 3.2: Attainment variables used to implement the CVA models

Variable	Type of variable	Description
Attainment in Mathematics	Outcome (Continuous standardised)	Mathematics test scores obtained by Year 10 pupils in 2006. Standardised scores range from -2.54 to 2.61. Original scores are estimated by the Ministry of Education using a three parameter logistic Item-Response Theory (3PL-IRT) model; scores range approximately from 100 to 400, with a mean of 250 and a standard deviation of 50.
Attainment in Spanish Language	Outcome (Continuous standardised)	Spanish Language test scores obtained by Year 10 pupils in 2006. Standardised scores range from -2.67 to 3.78. Original scores are estimated by the Ministry of Education using a three parameter logistic Item-Response Theory (3PL-IRT) model; scores range approximately from 100 to 400, with a mean of 250 and a standard deviation of 50.
Prior attainment in Mathematics	Explanatory (Continuous standardised)	Mathematics test scores obtained by Year 8 pupils in 2004. Standardised scores range from -2.96 to 2.93. Original scores are estimated by the Ministry of Education using a three parameter logistic Item-Response Theory (3PL-IRT) model; scores range approximately from 100 to 400, with a mean 250 and a standard deviation of 50.
Prior attainment in Spanish Language	Explanatory (Continuous standardised)	Spanish Language test scores obtained by Year 8 pupils in 2004. Standardised scores range from -3.36 to 2.62. Original scores are estimated by the Ministry of Education using a three parameter logistic Item-Response Theory (3PL-IRT) model; scores range approximately from 100 to 400, with a mean 250 and a standard deviation of 50.

Table 3.3: Explanatory variables at the pupil level used to implement the CVA models

Pupil-level Variables	Description and categories	Frequency (%)
Gender (categorical)	Gender of pupils:	
	Female (0)	103,496 (51.08%)
	Male (1)	95,916 (47.34%)
	Missing (.)	3,193 (1.58%)
Socio-economic level indicated by household income (categorical)	Average monthly household income in Chilean pesos (CLP) as reported by parents✝:	
	Low income (200,000 CLP or less) – reference category	99,939 (49.33%)
	Lower-middle income (200,001 CLP - 500,000 CLP)	54,086 (26.70%)
	Upper-middle income (500,001 CLP - 1,000,000 CLP)	18,147 (8.96%)
	High income (more than 1,000,001 CLP)	13,908 (6.86%)
	Missing (.)	16,525 (8.16%)
Year repetition (categorical)	It indicates whether the pupil has been made to repeat any year in primary school:	
	Not repeated (0)	171,014 (84.41%)
	Repeated (1)	18,956 (9.36%)
	Missing (.)	12,635 (6.24%)

✝ The parents' data were gathered by schools and sent as part of the required data for government auditing purposes. This thesis links the student, school and parental data.

Table 3.4: Explanatory variables at the secondary school level used to implement the CVA models

School-level Variables	Description and categories	Pupil Frequency (%)	School Frequency (%)
School institutional type (categorical)	State-funded schools (1) – reference category	83,127 (41.03%)	673 (27.55%)
	Subsidised independent schools (2)	103,819 (51.24%)	1,393 (57.02%)
	Independent schools (3)	15,659 (7.73%)	377 (15.43%)
	Missing (.)	0	0
School SES (categorical)	School Socio-economic status according to the classification of the Chilean Ministry of Education:		
	Low School SES (1) – reference category	36,035 (17.79%)	479 (19.61%)
	Lower-middle SES (2)	79,459 (39.22%)	665 (27.22%)
	Middle SES (3)	50,505 (24.93%)	584 (23.91%)
	Upper-middle SES (4)	21,942 (10.83%)	383 (15.68%)
	High SES (5)	14,664 (7.24%)	332 (13.59%)
	Missing (.)	0	0

Table 3.5: Pupil-level variables used to implement the measurement model for the latent construct of cultural capital

Variable†	Description	Frequency (%)
Parents' qualifications (categorical)	Highest level of education achieved by either the mother of the father of the pupil:	
	Primary school or less (1) - reference category	36,263 (17.90%)
	Secondary education (2)	92,570 (45.69%)
	Vocational education (3)	26,548 (13.10%)
	University degree or postgraduate (4)	37,164 (18.34%)
	Missing (.)	10,060 (4.97%)
Number of books in the household (categorical)	Estimated number of books in the household as reported by parents†:	
	No books - reference category (1)	2,391 (1.18%)
	Between 1 and 10 (2)	41,379 (20.42%)
	Between 11 and 50 (3)	82,016 (40.48%)
	Between 51 and 100 (4)	37,094 (18.31%)
	More than 100 (5)	38,851 (17.70%)
	Missing (.)	3,874 (1.91%)
Pupils' reading frequency (categorical)	Frequency with which pupils use reading materials as reported by parents†:	
	Pupil has no books - reference category (1)	19,330 (9.54%)
	Never (2)	11,663 (5.76%)
	Rarely (3)	61,301 (30.26%)
	Sometimes (4)	40,679 (20.08%)
	Always (5)	18,796 (9.28%)
	Missing (.)	50,836 (25.09%)
Parents' reading frequency (categorical)	Frequency with which parents read for pleasure as reported by parents†:	
	Never - Reference category (1)	14,548 (7.18%)
	Sometimes every month (2)	42,709 (21.08%)
	Sometimes every week (3)	64,497 (31.83%)
	Daily (4)	38,210 (18.86%)
	Missing (.)	42,641 (21.05%)

† These variables correspond to the parents' data that were gathered by schools and sent as part of the required data for government auditing purposes.

Table 3.6: Level identifier codes used to implement the CVA models

Level identifiers	Description
Pupil-level	Numerical code from 1 to N, based on the anonymous identifiers provided by the Chilean Ministry of Education, which allows linking pupils' data longitudinally. Number of pupils: 202,605.
Classroom	Numerical code provided by the Ministry of Education, identifying classrooms within schools cross-sectionally. This identifier does not allow to link classrooms longitudinally. Number of classrooms (within secondary schools): 7,487.
School	Numerical code from 1 to N, based on the publicly available school identifiers used by the Chilean Ministry of Education, which allows linking schools' data longitudinally, regardless of the educational level imparted. Primary and secondary schools are pooled to produce this code. Number of secondary schools: 2,443. Number of primary schools: 5,537.
Local Authority	Numerical code from 1 to N, based on the alphabetically ordered list of Local Authorities where schools are located. Number of local authorities: 320.

In the previous tables, missing data have been reported in many variables to varying degrees. The issues around handling missing data are discussed later in sections 3.7.3, 8.5, and appendix 4.

3.3. Measuring school effectiveness from a regression-based approach

Measuring school effectiveness traditionally involves analysing scores in standardised tests, controlling for a number of theorised influential covariates, such as demographic variables, socio-economic background, school characteristics, etc. Normally, one could approach this problem by using Multiple Regression Models in a standard and rather 'naive' fashion, which is analysing the level of linear association of this set of influential covariates with a relevant educational outcome, such as the scores in a standardised test of Mathematics or Language. Thus, one could have a number of coefficients for each variable defined to be significantly influential on the outcome, which would allow explaining a proportion of the recorded total variance, analysing the relative importance of each variable and making predictions based on the estimated equation.

The model in which a set of 'x' explanatory variables are used to explain a 'y' outcome variable is called the Multiple Linear Regression Model (MLR) (for more details, see for example: Dobson 2002). This model has the following underlying assumptions (Snijders & Bosker, 2011):

a) Observations are independent;
b) Error terms are independent;
c) Variance is constant across observations (homoscedasticity); and
d) Errors are normally distributed.

Meeting these requirements is critical for the model to work properly and to draw appropriate conclusions. The problem in the case of school effectiveness models is that the units of analysis are pupils who are nested within schools (other higher-level groupings can be defined, as discussed later), and hence, not independent, but clustered and thus correlated. Pupils from the same school tend to have more similar results than those from different schools, due to the selection processes put in place by schools and the shared environment (De Leeuw & Meijer, 2008; Hox, 2010). This is certainly a violation of the first assumption of the MLR approach (bullet point a).

As a consequence of this clustering, residuals derived from an MLR model would not be independent. The fact that pupils are not randomly assigned to their schools has a great impact on the error terms, which cannot be assumed to be random, but rather biased if nesting is not taken into account (De Leuuw & Meijer, 2008). Additionally, pupils in different schools may have errors with different variances. This is also known as heteroscedasticity and is a consequence of the error terms not being random or not explicitly allowing different residual variances at the higher-level units (De Leuuw & Meijer, 2008).

In sum, the standard MLR models are not robust to the violation of the assumption of independence of the observations (De Leeuw & Meijer, 2008; Hox, 2010; Snijders & Bosker, 2011) and thus clustering in the data to be analysed must not be ignored. Of substantive relevance is the fact that ignoring the hierarchical structure of the data leads to a large underestimation of the standard errors, which in turn would potentially result in the emergence of spuriously significant regression coefficients (De Leeuw & Meijer, 2008; Hox, 2010; Snijders & Bosker, 2011). The multilevel modelling framework, which allows a more thorough and reliable way of analysing clustered data, is described and discussed in more detail in the next section.

3.4. The multilevel modelling methodological framework

Multilevel modelling is an umbrella term for a wide range of statistical models appropriate for clustered data. Multilevel modelling can be thought of as an extension of the classical Multiple Regression Models (Hox, 2010) that allows the researcher to assess the variation in an outcome variable at different levels of a predefined hierarchy structure. The specification of multiple levels in the model results in the estimation of additional residual terms for every level defined in the model, which can also be correlated (if allowed for) to one or more explanatory variables (if any), but not to each other.

According to the particular way in which a multilevel model is specified, it can be named in different ways. For instance, a random intercepts model allows the intercept to vary at different levels, but specifies explanatory variables as fixed effects only (De Leeuw & Meijer, 2008; Hox, 2010; Snijders & Bosker, 2011). On another front, a random coefficients model (also known as random slopes model) can be understood as a further specification of the random intercepts model, which not only allows the intercept to vary at different levels, but also specifies additional error terms to allow the effect of one or more explanatory variables to vary randomly across multiple levels (De Leeuw & Meijer, 2008; Hox, 2010; Snijders & Bosker, 2011).

To illustrate how a multilevel model is constructed, one can start by adopting the somewhat 'naive' (for this purpose) MLR approach outlined in the previous section, where a basic single-level model is implemented to further build up from it. Following Goldstein's (2011) notation, a single-level regression model can be written as follows:

$$
\begin{gathered}
y_i = \beta_0 + (\mathrm{X\beta})_i + e_i \\
\textit{where:} \\
e_i \sim N(0, \sigma_e^2)
\end{gathered}
\tag{1}
$$

In equation 1, y_i is the dependent variable measured at the level of the individuals, β_0 is the intercept or average across individuals, $\mathrm{X\beta}$ is a vector of explanatory variables measured at the individual level along with their corresponding regression coefficients, and e_i is the error term, which is assumed to be normally-distributed and uncorrelated with the predictors. Due to the reasons outlined in the previous sections, the simplicity of this model is often considered to be unrealistic in the Social Sciences in general and educational research specifically, where many research problems involve dealing with hierarchically structured data (Hox, 2010; Snijders & Bosker, 2011).

Although the set of explanatory variables can include dummy variables as fixed effects for the relevant groups that produce the differences at the individual level, the adoption of this approach is discouraged by many researchers for various reasons (See for example: Rasbash, Steele, et al. 2012), including model parsimony, computational efficiency, as well as the impossibility of making inferences to the wider population.

A more complex and realistic model for dealing with nested data specifies an additional error term that describes the variability at the higher level(s) of the structure. The following equation describes the Variance Components Model:

$$
\begin{gathered}
y_{ij} = \beta_0 + u_{0j} + e_{ij} \\
\textit{or in multiple equations notation:} \\
y_{ij} = \beta_{0ij} \\
\beta_{0ij} = \beta_0 + u_{0j} + e_{ij} \\
\textit{where:} \\
u_{0j} \sim N(0, \sigma_u^2),\ e_{ij} \sim N(0, \sigma_e^2), cov(u_{0j}, e_{ij}) = 0
\end{gathered}
\tag{2}
$$

This model decomposes the total variation into two uncorrelated and normally distributed random effects at the level of individuals and at the level of groups. This model is useful as a baseline for assessing the relative importance of the multiple sources of variation and the level of clustering of the observations. This variance components model is then extended to specify explicitly explanatory variables, as equation 3 shows.

$$y_{ij} = \beta_{0ij} + (\mathrm{X}\beta)_{ij} + (\mathrm{X}\beta)_j + u_{0j} + e_{ij}$$

or in multiple equations notation:

$$y_{ij} = \beta_{0ij} + (\mathrm{X}\beta)_{ij} + (\mathrm{X}\beta)_j$$
$$\beta_{0ij} = \beta_0 + u_{0j} + e_{ij} \tag{3}$$

where:

$$u_{0j} \sim N(0, \sigma_u^2),\ e_{ij} \sim N(0, \sigma_e^2), cov(u_{0j}, e_{ij}) = 0$$

In equation 3, $(\mathrm{X}\beta)_{ij}$ and $(\mathrm{X}\beta)_j$ are vectors of explanatory variables specified at both levels of the structure (in this simple 2-level example model) and their corresponding regression coefficients. $(\mathrm{X}\beta)_j$ is suppressed from subsequent examples, because it is not central for the argument; it is only specified here to clarify that explanatory variables can be present at any level of the structure.

This equation represents the basic 2-level random intercepts model, in which two sources of variation are estimated in addition to the relationship between the explanatory variables (x_{ij}) and the outcome (y_{ij}). As a result, different regression lines are produced for each higher-level unit. This model assumes that the effects of the explanatory variables are constant across the groups of the sample and hence, the regression lines are parallel. In other words, the model assumes homoscedasticity at the higher-level units, which implies that the variation of the explanatory variables is constant across the higher-level groups.

Since this assumption is frequently regarded as unrealistic (De Leuuw & Meijer, 2008; Rasbash et al., 2009), the model can be extended further to allow the slopes to be random, as specified in equation 4.

$$y_{ij} = \beta_0 + \beta_{1j}x^*_{1ij} + u_{1j}x^*_{1ij} + (\mathrm{X\beta})_{ij} + u_{0j} + e_{ij}$$

or in multiple equations notation:

$$y_{ij} = \beta_{0ij} + \beta_{1j}x^*_{1ij} + (\mathrm{X\beta})_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{ij}$$

$$\beta_{1ij} = \beta_1 + u_{1j} \qquad (4)$$

where:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{0u} & \sigma_{01u} \\ \sigma_{01u} & \sigma^2_{1u} \end{pmatrix}\right)$$

$$e_{ij} \sim N(0, \sigma^2_{0e})$$

In equation 4, x^*_{1ij} is the explanatory variable that is allowed to vary randomly across the higher-level units, u_{1j} is the error term for the slope of the different groups and u_{0j} is the overall error term at the higher level. In this 2-level random coefficients model, a new error term has been added to account for differences in the variation between the groups of the sample. Thus, regression lines with different slopes can be drawn. This model is particularly useful in school effectiveness research, as the approach potentially allows for any effect to vary across schools (De Leuuw & Meijer, 2008), and thus relaxing the assumption of homoscedasticity.

As will be shown in the next section, when constructing a multilevel model for analysing school value-added, this basic 2-level random coefficients model can be extended to include more complex fully-nested or non-nested hierarchical structures, as well as more than one dependent variable.

3.5. School value-added in the multilevel modelling framework

Multilevel models have been used to measure the progress that pupils make between different stages during their schooling life. Multilevel models estimations are also used to produce school performance league tables for the purpose of public accountability, as well as informing parents' choice (Goldstein & Leckie, 2008; Goldstein & Spiegelhalter, 1996; Goldstein & Thomas, 1996; Leckie & Goldstein, 2011a). This last purpose is debatable as argued by Leckie and Goldstein (Goldstein & Leckie, 2008; Leckie & Goldstein, 2009), since the estimations proceeding from multilevel models, although retrospectively informative, do not seem to be reliable predictions of future performance, which is crucial for the purpose of school choice as discussed in chapter 2[9].

As mentioned earlier, the first step in the implementation of a multilevel model implies estimating the variance components. Recalling equation 2, the operational meaning of each term is summarised in Table 3.1.

The variance components model (equation 2) estimates the grand mean β_0 as well as the school effects u_{0j} and the pupil residuals e_{ij}. When school value-added is to be estimated, the focus of the analysis is set at the school level residuals, which represent the effect that a school has on the performance of its pupils. However, as this is an unconditional model (as defined by Plewis, 1996), these school effects cannot be understood as 'true' value-added estimates (Goldstein, 1997).

Table 3.7: Operational meaning of the terms of a 2-level variance components model for school effectiveness

Term	Operational meaning
y_{ij}	Pupils' scores in a standardised test
β_0	Overall (National) mean
u_{0j}	School effects
e_{ij}	Pupils' residuals
$y_{ij} = \beta_0 + u_{0j} + e_{ij}$	**Full equation (2-level variance components)**

Recalling equation 4, if x^*_{1ij} is assumed to be a measure of prior attainment, then the school effects u_{0j} can be regarded as value-added estimates, since what is being modelled under this framework is the change in attainment scores. Prior attainment can be therefore defined as a lagged outcome, which is denoted by $y_{t-1(ij)}$. Controlling for this lagged outcome is equivalent to model change directly by subtracting $y_{t(ij)} - y_{t-1(ij)}$. This can be more easily appreciated when equation 4 (but omitting the vector of explanatory variables for simplicity) is rearranged and rewritten in the following way (as seen on: Berrington et al. 2006; Plewis 1996):

$$y_{t(ij)} = \beta_0 + \beta_{1j}x^*_{1ij} + u_j + e_{ij}$$

Subtracting x^*_{1ij} in both sides of the equation, it gives:

$$y_{t(ij)} - x^*_{1ij} = \beta_0 + (\beta_1 - 1)x^*_{1ij} + u_j + e_{ij} \qquad (5)$$

and replacing $x^*_{1ij} = y_{t-1(ij)}$, it gives:

$$y_{t(ij)} - y_{t-1(ij)} = \beta_0 + (\beta_1 - 1)y_{t-1(ij)} + u_j + e_{ij}$$

[9] The issue of predicting future school performance is not central to this research, but it will be discussed briefly in later chapters.

In equation 5, pupils' progress, or change in attainment scores to be more precise, is a function of the overall mean and two uncorrelated normally-distributed random effects. An additional assumption is that the effect of prior attainment is fixed, which implies that it does not vary across schools. This value-added model constitutes a conditional (Plewis, 1996) or first order autoregressive (Steele, 2008) model. However, this model is still highly misfitting, misspecified and misleading as shown by some researchers (Ferrão & Goldstein, 2009; Foley & Goldstein, 2012; Leckie & Goldstein, 2009, 2011b). Researchers have found high levels of unobserved heterogeneity (Steele, 2008) in the relationship between prior and subsequent attainment, which is most likely the result of mutual dependence and time-invariant omitted variables (Steele, 2008).

These statistical issues have led researchers to argue in favour of what has come to be known as "contextualised value-added models" (Foley & Goldstein, 2012; Ray, 2006), which control this model misfit by including socio-economic, demographic and educational characteristics of pupils and schools. Rearranging equation 5 as a conditional 2-level model, controlling for prior attainment, pupil characteristics and school context, the resulting equation is:

$$\begin{gathered} y_{t(ij)} = \beta_{0ij} + \beta_1 y_{t-1(ij)} + (\mathrm{X}\beta)_{ij} \\ \beta_{0ij} = \beta_0 + u_j + e_{ij} \\ \textit{where:} \\ u_j \sim N(0, \sigma_u^2) \\ e_{ij} \sim N(0, \sigma_e^2) \end{gathered} \tag{6}$$

In equation 6, X is a vector of explanatory variables at the pupil and/or the school level. This is the most basic and common approach used for assessing contextualised school value-added. However, more recently, researchers (Ferrão & Goldstein, 2009; Fielding & Goldstein, 2006; Fielding et al., 2006; Foley & Goldstein, 2012; Leckie & Goldstein, 2009, 2011b; Leckie et al., 2010; Leckie, 2009) have found that CVA models do significantly benefit from the extension to more complex random effects structures, including additional random coefficients at the higher levels to control for heteroscedasticity, as well as additional levels of variation, which can also be non-fully hierarchical. Extending equation 6 to include additional higher levels and random coefficients for prior attainment, a 4-level random coefficients CVA model can be written as follows:

$$y_{t(ijkl)} = \beta_{0ijkl} + \beta_{1jkl} y_{t-1(ijkl)} + (\mathrm{X}\beta)_{ijkl}$$

$$\beta_{0ijkl} = \beta_0 + f_{0l} + v_{0kl} + u_{0jkl} + e_{0ijkl}$$

$$\beta_{1jkl} = \beta_1 + f_{1l} + v_{1kl} + u_{1jkl}$$

where:

$$\begin{pmatrix} f_{0l} \\ f_{1l} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{0f}^2 & \sigma_{01f} \\ \sigma_{01f} & \sigma_{1f}^2 \end{pmatrix} \right) \tag{7}$$

$$\begin{pmatrix} v_{0kl} \\ v_{1kl} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{0v}^2 & \sigma_{01v} \\ \sigma_{01v} & \sigma_{1v}^2 \end{pmatrix} \right)$$

$$\begin{pmatrix} u_{0jkl} \\ u_{1jkl} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{0u}^2 & \sigma_{01u} \\ \sigma_{01u} & \sigma_{1u}^2 \end{pmatrix} \right)$$

$$e_{0ijkl} \sim N(0, \sigma_{0e}^2)$$

This extension of the basic 2-level model of pupils nested within schools, can introduce for instance the levels of classrooms and local education authorities, as it has been done in other studies (Cervini, 2009a, 2009b; Goldstein et al., 2007; Leckie, 2009; Martínez, 2012; Plewis, 2011; Rasbash et al., 2010). Nevertheless, other higher levels can be specified, such as the teacher level and the province or region level. Furthermore, other random coefficients can also be specified to control for heteroscedasticity arising from other pupil-level variables; this can be useful to explore differential school effects for certain groups, usually defined by demographic characteristics such as gender and ethnicity (Kyriakides, 2004; Plewis, 2011).

Equation 7 is the algebraic form of the main model to be fitted in chapter 4 for progress in Mathematics and Spanish Language[10]. Both subjects are specified as outcomes in two separate univariate multilevel models. It needs to be pointed out that this is still a fully hierarchical model and, even though its complexity makes it a better approximation to the complex network of effects on pupils' performance than the traditional approach (2-level CVA models), it is still incomplete as it assumes that primary schools have no influence on pupils' performance and that their performance in different subjects are not related to each other. The value of such a model (a 4-level CVA model), as it will be discussed in more detail in chapter 4, lies in the fact that it is still informative of school effects in a more thorough way when compared to the traditional 2-level CVA models and this is done in a way that is fairly straightforward for a non-statistical audience.

[10] The main focus of Chapter 4 is set on the 4-level CVA models following the form of equation 7. Nevertheless, by the end of this chapter, an empty cross-classified model is fitted with the purpose of illustrating that there are further adjustments that can be made, but not as straightforwardly. In Chapter 5, this cross-classified specification is fully implemented.

In spite of its practical value, the 4-level CVA model controlling for the effects of local authorities and classrooms within secondary schools, makes some untenable assumptions from a statistical point of view. Considering that often pupils move from one school or neighbourhood to another during their educational trajectories, and that these conditions are likely to have differential effects on their performance, the model outlined in equation 7 can be extended and rearranged to allow for cross-classified effects as follows:

$$\begin{gathered} y_{t[i(j_1 j_2 j_3 j_4)]} = \beta_{0i(j_1 j_2 j_3 j_4)} + \beta_1 y_{t-1[i(j_1 j_2 j_3 j_4)]} + (\mathrm{X}\beta)_{i(j_1 j_2 j_3 j_4)} \\ \beta_{0i(j_1 j_2 j_3 j_4)} = \beta_0 + u_{j_1}^{(1)} + u_{j_2}^{(2)} + u_{j_3}^{(3)} + u_{j_4}^{(4)} + e_{i(j_1 j_2 j_3 j_4)} \end{gathered} \tag{8}$$

Note that this model does not contain a random coefficient for prior attainment and the variance structure of the random part has been omitted, for the sake of simplicity, although it follows the same pattern as shown in equation 7. It can also be appreciated that the subscripts are now all denoted by "j", because the structure is not strictly hierarchical. Specifying an additional random classification for primary school (or neighbourhood as noted earlier) makes the structure follow a cross-classified pattern, where, for instance, two pupils may have attended the same primary school, but not necessarily the same secondary school. Such a structure sets all the higher random classifications at the same level, and hence the "j" subscript for all. The multiple subscript notation in equation 8 may be deemed as excessive, and thus it is rearranged as follows:

$$\begin{gathered} y_{t(i)} = \beta_{0i} + \beta_1 y_{t-1(i)} + (\mathrm{X}\beta)_i \\ \beta_{0i} = \beta_0 + u_{0,lea(i)}^{(1)} + u_{0,second(i)}^{(2)} + u_{0,prim(i)}^{(3)} + u_{0,class(i)}^{(4)} + e_i \end{gathered} \tag{9}$$

Once again, the model has been intentionally simplified, so it contains no random coefficients for the pupil-level variables of the fixed part; moreover, the variance structure has also been omitted, since it follows the same structure as outlined in equation 7. This is a cross-classified 5-level model (Fielding & Goldstein, 2006; Fielding et al., 2006; Goldstein et al., 2007; Hill & Goldstein, 1998; Simonite & Browne, 2003), where full nesting of levels is not assumed. For this reason, note that the model in equation 9 no longer identifies the higher-level residuals with different subscripts ("j", "k", "l" and so forth), and now they are all denoted by an abbreviation of the level they correspond to and the "i" subscript for pupils. The superscripts in parentheses for the higher-level residuals are used to denote the number of each level (this is especially useful for the random part of the model). The levels specified in this model are the same as in equations (8) and (9), namely: local education authorities (lea); secondary schools (second); primary schools (prim); classrooms (class) and pupils (i).

This cross-classified structure does not have the unrealistic underlying assumption that the only school that has an influence on the outcome (specified at the pupil level) is the secondary school to which the pupils has attended. Thus, it does not ignore the primary school. As mentioned before, this has proved to be particularly useful in educational research, where it is relevant to estimate the carry-over effects of primary school into secondary school performance (Leckie, 2009; Rasbash et al., 2010).

Despite its more realistic complexity, the specification of additional levels of variation in this extended CVA model is still incomplete. Up to this point, only one dependent variable has been specified in the CVA models; however, further extensions can also include multivariate response multilevel models, where two or more dependent variables and their corresponding relationships can be analysed. Following Goldstein's (2011) notation, the most basic 2-level multivariate model can be written as in equation 10 below.

$$
\begin{gathered}
y_{ij} = \beta_{01ij}z_{1ij} + \beta_{02ij}z_{2ij} + (\mathrm{X}\beta)_{1ij}z_{1ij} + (\mathrm{X}\beta)_{2ij}z_{2ij} \\
\beta_{01ij} = \beta_{01} + u_{1j} + e_{1ij} \\
\beta_{02ij} = \beta_{01} + u_{2j} + e_{2ij} \\
\textit{where:} \\
z_{1ij} = \begin{pmatrix} 1\ if\ outcome\ 1 \\ 0\ if\ outcome\ 2 \end{pmatrix}, z_{2ij} = 1 - z_{1ij} \\
\textit{and:} \\
\begin{pmatrix} u_{1j} \\ u_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u12} & \sigma_{u2}^2 \end{pmatrix} \right) \\
\begin{pmatrix} e_{1j} \\ e_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{e1}^2 & \sigma_{e12} \\ \sigma_{e12} & \sigma_{e2}^2 \end{pmatrix} \right)
\end{gathered}
\tag{10}
$$

In equation 10, the bivariate outcome y_{ij} is a twofold set of outcome variables defined by the dummy variables z_{1ij} and z_{2ij}. This model fits two equations simultaneously. The data set no longer has the wide format as in previous models; it now has the long format with two observations per case (one for each outcome). A more detailed example is appreciated in Table 3.8. Note that in equation 10, prior attainment has been suppressed and merged into the vector of explanatory variables X, as this makes no difference in the estimation. It can also be appreciated that the variance structure of the random part is more complex than in previous models, since covariance terms between the random intercepts of both outcomes are to be estimated at all levels; this is the element that links both equations fitted for z_{1ij} and z_{2ij}. Obviously, this basic bivariate 2-level model can be extended to situations where residual slope variances and additional levels are to be estimated.

Table 3.8: Example of data set in the long format for bivariate model

Pupil	Response	Outcome (1) z_{1ij}	Outcome (2) z_{2ij}
1	$y_1^{(1)}$	1	0
1	$y_1^{(2)}$	0	1
2	$y_2^{(1)}$	1	0
2	$y_2^{(2)}$	0	1
3	$y_3^{(1)}$	1	0
3	$y_3^{(2)}$	0	1

Source: Adapted from Goldstein (2011)

The following equation is an extension of equation 10 for the case of a model with two dependent variables, which can be for instance standardised tests scores of two core subjects, and five non-fully hierarchical levels. For this study, the two outcome variables are Mathematics and Spanish Language scores in the standardised SIMCE tests.

$$y_i = \beta_{01} z_{1i} + \beta_{02} z_{2i} + (\mathrm{X\beta})_{1i} z_{1i} + (\mathrm{X\beta})_{2i} z_{2i}$$

$$\beta_{01} = \beta_{01i} + u_{1,lea(i)}^{(1)} + u_{1,second(i)}^{(2)} + u_{1,prim(i)}^{(3)} + u_{1,class(i)}^{(4)} + e_{1i}$$

$$\beta_{02} = \beta_{02i} + u_{2,lea(i)}^{(1)} + u_{2,second(i)}^{(2)} + u_{2,prim(i)}^{(3)} + u_{2,class(i)}^{(4)} + e_{2i}$$

where: (11)

$$z_{1i} = \begin{pmatrix} 1 \; if \; outcome1 \\ 0 \; if \; outcome2 \end{pmatrix}, z_{2i} = 1 - z_{1i}$$

and:

$$\begin{pmatrix} u_{1,lea(i)}^{(1)} \\ u_{2,lea(i)}^{(1)} \end{pmatrix} \sim N\left(0, \Omega_{u^{(1)}}\right), \begin{pmatrix} u_{1,second(i)}^{(2)} \\ u_{2,second(i)}^{(2)} \end{pmatrix} \sim N\left(0, \Omega_{u^{(2)}}\right),$$

$$\begin{pmatrix} u_{1,prim(i)}^{(3)} \\ u_{2,prim(i)}^{(3)} \end{pmatrix} \sim N\left(0, \Omega_{u^{(3)}}\right), \begin{pmatrix} u_{1,class(i)}^{(4)} \\ u_{2,class(i)}^{(4)} \end{pmatrix} \sim N\left(0, \Omega_{u^{(4)}}\right),$$

$$\begin{pmatrix} e_{1i} \\ e_{2i} \end{pmatrix} \sim N(0, \Omega_e)$$

In equation 11, all Ω's follow the variance-covariance structure of equation (10) and prior observations of both outcomes (prior attainment in the corresponding subject) have been merged into their corresponding vector of explanatory variables X. Furthermore, the levels specified in this model are the same as in equation 9. This model does not specify random coefficients for any explanatory variable, and hence, it can be regarded as a 5-level random intercepts bivariate cross-classified CVA model. Nevertheless, it can still be further specified to include random coefficients for the explanatory variables to test for differential school effects.

When random coefficients are specified, the model increases in complexity notably, because of the number of additional random-effects parameters at all the higher levels, were they fully specified. In this research, prior attainment in Mathematics and Spanish Language have been specified as random effects only at the level of secondary schools, since the focus is set at the school level; however, the model would eventually allow for more random effects at the other higher levels of the structure. As before, following the general multilevel notation as described in Goldstein (2011) and the classification notation described in Browne et al. (2001), the random coefficients bivariate cross-classified 5-level model to be fitted in chapter 5 has the algebraic form of equation 12.

$$
\begin{gathered}
y_{t(i)} = \beta_{01}z_{1i} + \beta_{02}z_{2i} + \beta_{11}z_{1i}y_{t-1(i)} + \beta_{12}z_{2i}y_{t-1(i)} + (X\beta)_{1i}z_{1i} + (X\beta)_{2i}z_{2i} \\
\beta_{01} = \beta_{01i} + u^{(1)}_{1,local(i)} + u^{(2)}_{1,secondary(i)} + u^{(3)}_{1,primary(i)} + u^{(4)}_{1,classroom(i)} + e_{1i} \\
\beta_{02} = \beta_{02i} + u^{(1)}_{2,local(i)} + u^{(2)}_{2,secondary(i)} + u^{(3)}_{2,primary(i)} + u^{(4)}_{2,classroom(i)} + e_{2i} \\
\beta_{11} = \beta_{11i} + u^{(2)}_{3,secondary(i)} \\
\beta_{12} = \beta_{12i} + u^{(2)}_{4,secondary(i)} \\
where \\
z_{1i} = \begin{Bmatrix} 1\ if\ Mathematics \\ 0\ if\ Language \end{Bmatrix}, z_{2i} = \begin{Bmatrix} 1\ if\ Language \\ 0\ if\ Mathematics \end{Bmatrix}, \\
and \\
\begin{bmatrix} u^{(1)}_{1,local(i)} \\ u^{(1)}_{2,local(i)} \end{bmatrix} \sim MVN\left(0, \Omega^{(1)}_u\right) : \Omega^{(1)}_u = \begin{bmatrix} \sigma^2_{u11} & \sigma_{u11,12} \\ \sigma_{u11,12} & \sigma^2_{u12} \end{bmatrix} \\
\begin{bmatrix} u^{(2)}_{1,secondary(i)} \\ u^{(2)}_{2,secondary(i)} \\ u^{(2)}_{3,secondary(i)} \\ u^{(2)}_{4,secondary(i)} \end{bmatrix} \sim MVN\left(0, \Omega^{(2)}_u\right) : \Omega^{(2)}_u = \begin{bmatrix} \sigma^2_{u21} & \sigma_{u21,22} & \sigma_{u21,23} & \sigma_{u21,24} \\ \sigma_{u21,22} & \sigma^2_{u22} & \sigma_{u22,23} & \sigma_{u22,24} \\ \sigma_{u21,23} & \sigma_{u22,23} & \sigma^2_{u23} & \sigma_{u23,24} \\ \sigma_{u21,24} & \sigma_{u22,24} & \sigma_{u23,24} & \sigma^2_{u24} \end{bmatrix} \\
\begin{bmatrix} u^{(3)}_{1,primary(i)} \\ u^{(3)}_{2,primary(i)} \end{bmatrix} \sim MVN\left(0, \Omega^{(3)}_u\right) : \Omega^{(3)}_u = \begin{bmatrix} \sigma^2_{u31} & \sigma_{u31,32} \\ \sigma_{u31,32} & \sigma^2_{u32} \end{bmatrix} \\
\begin{bmatrix} u^{(4)}_{1,classroom(i)} \\ u^{(4)}_{2,classroom(i)} \end{bmatrix} \sim MVN\left(0, \Omega^{(4)}_u\right) : \Omega^{(4)}_u = \begin{bmatrix} \sigma^2_{u41} & \sigma_{u41,42} \\ \sigma_{u41,42} & \sigma^2_{u42} \end{bmatrix} \\
\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \sim MVN\left(0, \Omega_e\right) : \Omega_e = \begin{bmatrix} \sigma^2_{e1} & \sigma_{e12} \\ \sigma_{e12} & \sigma^2_{e2} \end{bmatrix}
\end{gathered}
\tag{12}
$$

This model not only would allow for more random coefficients, as already mentioned, but also more outcome variables. Given that school performance can be understood as a multidimensional phenomenon, a more thorough approach would need to include more outcome variables, such as pupils' scores in Science, History and whatever other subject is assessed, along with non-educational outcomes such as motivation, school climate, satisfaction, etc. Regretfully, the 2004-2006 SIMCE databases are limited to the longitudinal assessment of pupils' attainment in Mathematics and Spanish Language and do not include any assessment of non-educational outcomes, and hence the model can only be specified as a bivariate multilevel model. Nonetheless, extending this model further would be rather straightforward, insofar as it would only require additional dummy variables ("z" indicators in equation 12) and random intercepts at all higher levels, which would follow the same variance-covariance structure already explained. Although, its ever-increasing complexity would make it even less accessible for a non-statistical audience.

3.6. Variance composition and residuals in multilevel models

Analysing the composition of the total variance by splitting it into diverse sources (levels) allows assessing the relative importance of each component/level. The variance partition coefficient (VPC) is an important diagnostic measure that allows to assess whether it is worthwhile to utilise a multilevel model or to keep a single-level model, since it provides an indication of how biased the estimates of a single-level model are expected to be (Goldstein, 2011; Hox, 2010; Snijders & Bosker, 2011). From a substantive point of view, in a school value-added model, the VPC allows to assess the relative "weight" of the school residuals with respect to all other sources of variation. Recalling equation (2), the 2-level variance components model assumes that the estimates of the residuals at the school (u_j) and pupil level (e_{ij}) are uncorrelated and follow a normal distribution with a mean of 0 and variances σ_u^2 and σ_e^2. The proportion of the total variance accounted for by the higher-level units can be assessed by computing the variance partition coefficient as follows:

$$\rho_u = \frac{\sigma_{0u}^2}{\sigma_{0u}^2 + \sigma_{0e}^2} \tag{13}$$

This coefficient is also known as 'intra-class correlation' (Snijders & Bosker, 2011) or 'intra-school correlation' in a basic 2-level school value-added model. In models with three and more fully-hierarchical levels, there are two ways of calculating the variance partition coefficient (Hox, 2010). The first way is expressed in equations 14a and 14b; this is an extension of the equation for a 2-level model (equation 13), where the proportion of the variation exclusively due to a single level is calculated, and hence, the numerator of the equation only includes the variance estimate of the particular level of interest, whereas the denominator is defined as the sum of the variances of all levels. For a 3-level model, this is expressed by the two forms that equation 14 can take.

$$\begin{aligned} &\text{a)} \qquad \rho_u = \frac{\sigma_{0u}^2}{\sigma_{0v}^2 + \sigma_{0u}^2 + \sigma_{0e}^2} \\ &\text{b)} \qquad \rho_v = \frac{\sigma_{0v}^2}{\sigma_{0v}^2 + \sigma_{0u}^2 + \sigma_{0e}^2} \end{aligned} \tag{14}$$

In equations 14a and 14b, the subscripts "v", "u" and "e" denote the residuals of levels 3, 2 and 1, respectively, which can be, for instance, the levels of secondary schools, classrooms and pupils. For the case of a 4-level model, equations 14a and 14b can be extended to include the variance of the intercept at the highest level, i.e. local authorities in this research. The fourth variance would be included in the denominator for calculating the VPC of each level of the structure and in the numerator to calculate the VPC of that particular additional higher level.

The second method takes into account the nested structure (Siddiqui et al., 1996). Thus in a 3-level fully-nested model, lower-level units that belong to the same cluster, also belong to the same super-cluster. Hence, the variances of both higher-level units are included in the numerator of the equation as follows:

$$\rho_u = \frac{\sigma_{0v}^2 + \sigma_{0u}^2}{\sigma_{0v}^2 + \sigma_{0u}^2 + \sigma_{0e}^2} \qquad (15)$$

This method can be understood as a 'combined VPC'. Both methods can be easily extended to situations where more fully-hierarchical levels are specified, by including the additional variance estimate in the denominator and the variance estimate of interest in the numerator. It should be noted though that the 'combined VPC' method is not applicable for models where full nesting of levels is not assumed, such as cross-classified models (model to be implemented in chapter 5), and hence only the 'exclusive VPC' approach must be adopted.

Despite its importance, the VPC does not give any estimation of the goodness of fit, which is best assessed by the likelihood ratio test for nested models, where the single-level model is used as a baseline (Goldstein, 2011; Hox, 2010). This method can be extended to situations where more fully-hierarchical levels are specified, using the last estimated model as the baseline (Goldstein, 2011; Hox, 2010).

Nevertheless the aforementioned, educational researchers have largely demonstrated that multilevel models are the current best approach for analysing pupil and schools' performance in standardised tests (Goldstein, 1997). For this reason, the basic 2-level model of pupils nested within schools is the starting point of the modelling process.

Since it has been determined that the multilevel structure does imply an improvement of the overall fit of the model, and hence it is the best approach, the focus of the modelling process is set on the estimation of the most reliable and unbiased school effects. The following equation shows how the schools' residual point estimates are computed in a basic 2-level model of pupils nested within schools:

$$\hat{u}_j = \frac{n_j \sigma_u^2}{\left(n_j \sigma_u^2 + \sigma_e^2\right)} \bar{r}_j$$

$$\textit{where:} \qquad (16)$$

$$\bar{r}_j = \frac{\sum\left(y_{ij} - \beta_0\right)}{n_j}$$

This is known as the posterior estimate of the school effects or 'shrunken residuals' (Goldstein, 2011); where n_j is the size of the j-th school; σ_u^2 is the between-school variance; and $\bar{r}_j$ is the mean of the pupils' raw residuals of the j-th school. These shrunken school residual estimates explicitly incorporate uncertainty by adjusting the effects according to the size of the schools and the variability recorded in the scores of their own pupils, which produces more conservative value-added estimates. The variance associated with these school effects is estimated as follows:

$$var(\hat{u}_j - u_j) = \frac{\sigma_u^2 \sigma_e^2}{n_j \sigma_u^2 + \sigma_e^2} \tag{17}$$

This variance can be regarded as a 'comparative variance' (Goldstein et al., 2002), as it allows comparing the raw school residuals with their 'true' values. Here, a 'superpopulation' perspective is adopted (for a wider discussion on the superpopulation models, see for example: Isaki & Fuller, 1982), because the main interest in this analysis is to establish the mean effect of this theorised underlying process that is generating these results, in order to produce better estimates (Goldstein et al., 2002). This variance can also be used to compute confidence intervals for the 'true' value of the school residuals. Thus, assuming that the schools effects are normally distributed, a 95% confidence interval for the effect of the j-th school can be estimated as shown in equation 18.

$$\hat{u}_j - 1.96\sqrt{var(\hat{u}_j - u_j)} < u_j < \hat{u}_j + 1.96\sqrt{var(\hat{u}_j - u_j)} \tag{18}$$

In equation 18, -1.96 and 1.96 are, respectively, the approximate values of the 2.5 and 97.5 percentiles of the standard normal distribution. This confidence interval might then be used to determine whether the effect of a particular school is significantly different from zero and produce more reliable school performance tables, which is the common practice in England (Foley & Goldstein, 2012; Ray, 2006).

Constructing 95% confidence intervals for the school residual point estimates with 1.96 standard deviations is useful for comparing schools with the overall average. However, Goldstein and Healy (1995) have shown that using 1.39 standard deviations (instead of the usual 1.96) for constructing the confidence intervals is especially useful for making pairwise school comparisons. This approach has not been implemented, as it escapes the scope of this research.

In this section, only the most basic form of the school value-added estimates has been described. As it will be appreciated in chapters 4 and 5, as the models become more complex, the school value-added estimates shrink further, due to the effect of the explanatory variables and the further specifications of the model, such as the inclusion of random coefficients. When these specifications are made, the traditional notions of 'school value-added' or even 'contextualised value-added' no longer apply, although the focus of the modelling process is still on the estimation of the school residuals.

3.7. Analytical strategy

3.7.1. Construction of the cultural capital indicator

In order to avoid possible problems of multicollinearity by adding different variables theoretically related to cultural capital and to each other, a latent variable approach was adopted to analyse the effect of cultural capital on pupils' progress in chapter 6. Four variables imported from the parents' survey can be thought of as realisations of the latent construct "cultural capital" (Bourdieu, 1986). These four variables are: a) parents' educational qualifications; b) number of books in the household; c) frequency with which pupils read "for pleasure" (fiction books, novels and/poetry, magazines, newspapers, etc.); and d) frequency with which parents read "for pleasure".

These four variables are ordinal, which is why the traditional approach of factor analysis cannot be implemented reliably. One of the appropriate statistical procedures to be implemented is confirmatory factor analysis for ordered categorical data (Hancock & Mueller, 2006; Kaplan, 2009; Kline, 2011; Muthén, 1978, 1983, 1984). This is also known in the literature as Item Response Theory (IRT) models (de Ayala, 2013; Kline, 2011). This analysis can be easily implemented in the software package Mplus (Muthén & Muthén, 2012). The model implemented contains one latent variable called "cultural capital" defined by the four ordered categorical measures mentioned above. The measurement model is a two-parameter probit model and is estimated using the WLSMV algorithm (Mean-and-Variance-adjusted Weighted Least Squares), which is the default estimator for categorical data in Mplus, because of its robustness demonstrated in simulation studies (Yu, 2002).

Several measures of goodness of fit were used to assess the measurement model since there is not any all-purpose fit measure. The Chi-Square test assesses how well the specified model fits the data, by judging the discrepancies between the sample and the model covariance matrices (Hooper et al., 2008). This is reported in this research; however, the large number of observations makes it unreliable for being overly sensitive to small departures from normality. Other measures used to assess the fit of the measurement model are those that assess the discrepancies between a baseline model and the measurement model; these are known as incremental, relative or comparative fit indices (Hooper et al., 2008): a) the Tucker-Lewis Index (TLI); b) the Comparative Fit Index (CFI). On the other hand, the Root Mean Square Error of Approximation (RMSEA) is a measure of the differences between the sample covariance matrix and the fitted covariance matrix and it has been regarded as one of the most informative and robust fit indices (Hooper et al., 2008). Finally, the last fit index used in this research is the Weighted Root Mean Square Residual (WRMR), which measures the weighted average residual differences between the estimated and the population variances and covariances (Yu, 2002). The criteria to judge the values of these measures were obtained from Hooper et al. (2008), Yu (2002), Kaplan (2009), Kline (2011), as well as Finney and DiStefano (2006).

From this measurement model, a predicted latent normally-distributed variable was derived to represent "cultural capital" as an individual-level continuous explanatory variable in the bivariate multilevel model to be implemented in chapter 6. The multilevel models presented in chapter 6 follow the same analytical strategy as for the models in Chapter 4 and 5, which is outlined in the next section.

3.7.2. Multilevel model building approach

The models implemented in chapters 4, 5 and 6 have been built sequentially, following a strategy that has been adapted from Hox (2010) to include the assessment of the addition of multiple levels to the model. The analyses have been performed by using the software Stata, version 12.1 (StataCorp, 2011) and the user-built Stata module "runmlwin" (Leckie & Charlton, 2013) which runs the software MLwiN (Rasbash, Charlton, et al., 2012) from within Stata. Each step uses the estimates stored from the previous model as the starting values for the parameters to estimate in the subsequent model.

The parameters of the models considering a fully-nested structure (chapters 4 and 6) have been estimated by using the Iterative Generalised Least Squares method (IGLS) which is the default estimation method for normally distributed variables in MLwiN. In the case of the cross-classified models of chapter 5, IGLS estimates are used as a way of providing reliable starting values for the Markov Chain Monte Carlo (MCMC) estimation. MCMC estimation, in the model exploration stage, was carried out with the Gibbs sampler using diffuse priors, a burn-in period of 500 iterations, a monitoring chain length of 5,000, storing all iterations (thinning equals to 1). For the full model in chapter 5, the monitoring chain length was increased to 215,000, storing all iterations. The necessity of treating the levels as cross-classified came to be when adding the level of primary schools.

Estimation via MCMC in MLwiN is needed, because it is more computationally efficient than fitting a cross-classified model using maximum likelihood (ML estimates, which are obtained via the IGLS algorithm). This is because obtaining ML estimates requires a series of constraints that increase the computational burden even in the univariate case. The bivariate case would require even further extensions to the estimation algorithms that are not readily available in other software packages, such as Stata or R. For instance, the "lme4" package in R does support ML estimation of cross-classified models; however, it does not handle multivariate outcomes. The same occurs for the "mixed" module in Stata. On the contrary, MLwiN has the capability of handling multiple outcomes and crossed effects simultaneously and it is fairly straightforward to implement using MCMC estimation.

The statistical significance of the specification of every level was judged either by using the Akaike Information Criterion (AIC) or the Likelihood Ratio Test (LR test) for the case of the models with fully-nested structures (chapters 4 and 6); and by using the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) for the case of the bivariate cross-classified models of chapter 5. A full description of these tests can be found in Snijders and Bosker (2011), Goldstein (2011) and Browne (2012). In sum, the analytical strategy followed in this research can be summarised in these seven steps:

Step 1: "Analyse a model with no explanatory variables" (Hox, 2010, p. 56). This step starts by fitting a 2-level variance component model for attainment in Mathematics or Spanish Language (or both simultaneously in chapters 5 and 6) and set this as a benchmark for the next steps that involve specifying additional levels. Since there is evidence in previous research that the misspecification of the number of levels can lead to bias in the models (Tranmer & Steel, 2001; Van Landeghem et al., 2010), the specification of additional levels ought to be assessed as well. This step corresponds to what Timmermans et al. (2011) denote as "type 0 value-added" or "gross school effects". Within this step, the following sub-steps are also taken:

a) Analyse a 2-level empty model. This model is set as the baseline model with the most frequently analysed structure in educational research, i.e. pupils nested within (secondary) schools.
b) Analyse a 3-level empty model. Since it is theoretically and empirically plausible that the classrooms, to which pupils are assigned, are a source of variation in academic performance, a 3-level empty model is fitted. Its deviance is then compared to the baseline model to assess whether it is a significant addition to the model. This model is structured as pupils nested within classes within schools.
c) Analyse an alternative 3-level empty model. If the structure defined in the step 1b) is not found to be a significant addition to the model, an alternative 3-level structure may be assessed, which is pupils nested within schools and schools nested within local authorities. The inclusion of these as a higher-level in the analysis is performed given the empirical evidence of diverse levels of pupils' performance and high levels of socio-economic inequality amongst local authorities. This sub-step renders redundant if sub-step 1b yields significant results and is therefore skipped.

d) Analyse a 4-level empty model. Having tested and proved the significance of the model proposed in b), the next step is to assess the contribution of the addition of a fourth level. As mentioned in c), there is empirical evidence that suggests that there is a geographic component in pupils' performance. Thus, a 4-level model of pupils nested within classrooms within schools within local authorities is implemented.
e) Analyse a 5-level cross-classified empty model. Having tested the significance of the previous model (sub-step 1d), the next sub-step is to assess whether the additional non-nested level of primary schools significantly contributes to improve the overall fit of the model. This model specifies the level of primary school as an additional random effect; however, since pupils from the same secondary may have gone to different primary schools, this model is not completely hierarchical, which requires to fit a cross-classified multilevel model, where a not fully-nested structure is implied[11].

Step 2: An additional step has been included in this strategy to specify a conditional model to take into account the variance from the prior scores in Mathematics or Spanish Language. As mentioned previously, a true value-added model needs to control for prior attainment as this allows having a measure of progress from one stage to another. This also allows setting a baseline for estimating the amount of variance accounted for by the model at all levels. In this step, the variable "prior attainment" (either in Mathematics or Spanish Language separately, or simultaneously) is added to the model, which is the score that pupils obtained in the previous occasion of the test. This corresponds to a "type AA value-added" model (Timmermans et al., 2011).

Step 3: "Analyse a model with all lower-level explanatory variables fixed" (Hox, 2010, p. 56). In this step, the pupil-level variables are added, namely: gender, retention and average monthly household income. Furthermore, the following relevant interaction terms are specified: prior attainment and gender; prior attainment and income; prior attainment and retention; gender and income; gender and retention and; income and retention. This corresponds to the "type A value-added" (Timmermans et al., 2011), which would be the most useful for school choice purposes from a substantive perspective. Methodologically, this model can be thought of as an intermediate model that allows controlling for compositional effects, when the primary aim is to isolate the school effects.

[11] This model is implemented and presented at the end of chapter 4, only for the case of attainment in Mathematics, to show how carry-over effects from primary school are relevant sources of variation in academic performance. However, it has not been implemented with the full set of explanatory variables, since it is highly time-consuming and, more importantly, it is beyond the main aim of that chapter. Nevertheless, a cross-classified structure is fully implemented in the bivariate CVA models for progress in Mathematics and Spanish Language of chapter 5.

Step 4: "Add the higher-level explanatory variables" (Hox, 2010, p. 57). In this step, besides including the pupil-level variables from the previous step, the school-level variables (school's institutional type and school's average socio-economic status - SES) are added to the model, as well as the interaction between both variables. The intermediate model fitted in this step and all the forthcoming are all classified as "type X value-added" models (Timmermans et al., 2011), since school context and non-malleable factors are included in this step onwards.

Step 5: "Assess whether any of the slopes of any of the explanatory variables has a significant variance component between the groups" (Hox, 2010, p. 57). In this step, prior attainment scores and gender are allowed to vary randomly across all higher-levels. The significance of these random coefficients is assessed by performing the LR test and comparing the AIC values.

Step 6: "Add cross-level interactions between explanatory group-level variables and those individual explanatory variables that had significant slope variation" (Hox, 2010, p. 58). In this step, the interaction between prior attainment scores, gender and the school-level variables institutional type and school SES are added in the model. The significance of these interactions terms is assessed by performing the Wald test.

Step 7: Fitting a 2-level model with the same specifications as the final 4-level CVA model for comparison. This step is not included in Hox's strategy. This is performed with the purpose of demonstrating the practical usefulness and statistical superiority of the extended CVA models fitted in this research. These models are comparable insofar as they are nested, i.e. the 2-level CVA model can be thought of as a constrained 4-level CVA model, where the intercept is not allowed to vary randomly across the additional higher levels of classrooms and local authorities.

This analytical strategy was used to implement the univariate CVA models for progress in Mathematics and Spanish Language presented in chapter 4. It was also used to implement the bivariate CVA model presented in chapter 5, where instead of building two separate models for progress in Mathematics and Spanish Language, a model with two dependent variables was built taking into account the correlation between both subjects as seen in equation 12. The models presented in chapter 6 also follow this strategy, with the exception of step 7, which is irrelevant for the purpose of that chapter.

3.7.3. Dealing with missing data

Needless to say, handling missing data appropriately is very important in statistical analysis; however, this in itself is a complicated matter. For the case of the models in chapters 4 and 5, which use data from tables 3.2, 3.3, 3.4 and 3.6, the treatment of the missingness is listwise deletion. This is done for reasons of parsimony and convenience as in other previous studies (see for example: Leckie, 2009; Rasbash et al., 2010). The most appropriate way of handling this issue from a statistical point of view is multilevel multiple imputation; however, this also has its limitations. As discussed later in chapters 7 and 8, performing post-estimation procedures on multiply imputed datasets is particularly cumbersome (and thus usually avoided) in multilevel and latent variable settings (Gelman et al., 2005). Moreover, standard procedures are not readily implemented in any software package. For instance, MLwiN can indeed estimate models from multiply imputed datasets; however, it only estimates residuals for the complete cases. Even though a reasonable doubt can be cast on the results from the models using complete cases only, results from a multilevel multiple imputation procedure performed on a sample of the full dataset showed very little bias in the coefficients when comparing the results from a complete-case analysis and the pooled results from the multiply imputed datasets. The multilevel multiple imputation models were run using the software package Realcom-Impute (Carpenter et al., 2011) in combination with MLwiN. Full details are given in appendix 4.

In the case of the measurement models of chapter 6, which use the data presented in Table 3.5, the handling of missing data was pairwise deletion. This is implemented by default in the software package Mplus (Muthén & Muthén, 2012), in which the measurement models were run. Pairwise deletion in the Mean and Variance Adjusted Weighted Least Squares (WLSMV) algorithm, allows estimating the covariance matrix efficiently, even in the presence of missing data. The measurement model of chapter 6 is estimated with 199,114 cases, i.e. 98.27% of the total number of cases.

3.8. Ethical Issues and Considerations

The main ethical issues that arise from this research are associated with the treatment of personal data, confidentiality and the risk of disclosure. Personal information referring to pupils, teachers, parents and carers in the SIMCE databases has been anonymised by the Chilean Ministry of Education. For the case of schools, an anonymisation procedure has been implemented to ensure that school identifiers are encoded in an unidentifiable and yet consistent way.

Despite these measures, the risk of disclosure is still an issue that needs to be tackled. For this matter, the recommendations from the UK Office for National Statistics (ONS, 2006) regarding the reduction of the risk of disclosure are to be followed. This implies the careful release of descriptive statistical tables, where no cell must contain four or less cases, because this produces a high probability of identification.

Chapter 4: Analysing value-added for external school accountability from a univariate perspective

4.1. Introduction

The main purpose of this chapter is firstly to analyse how school value-added and pupils' academic progress are affected by factors external to the widely known effects of the socio-economic and demographic characteristics of the pupils and the schools. As discussed in detail in previous chapters, the traditional approach to school effectiveness has involved the analysis and implementation of 2-level models, where pupils are nested within school; however, recent research has shown how these models, although still informative to a certain extent, can be insufficient to tackle the underlying complexity of academic performance.

In this chapter, a 4-level CVA model for progress in Mathematics is developed and discussed in detail in a step by step fashion, with the purpose of illustrating the whole modelling process towards a more reliable and fair model for external school accountability. In this extended model, the additional levels of classrooms and local authorities are specified and proved to be significant sources of variation in progress. Later on, the more complex extended specification of this CVA model is compared to a simpler specification akin to the traditional 2-level approach, with the purpose of demonstrating its practical usefulness and statistical superiority.

Afterwards, the 4-level model is extended even further to assess the carry-over effects of primary schools in Mathematics progress as a starting point for the analyses carried out in chapter 5. By the end of this chapter, an analogous extended 4-level CVA model is fitted with Spanish Language as the outcome and the differences with respect to what is recorded in the case of Mathematics are discussed in detail. The CVA models fitted throughout this chapter are univariate and deliberately so, because the purpose is to propose a relatively straightforward extension to the traditional approach that can be readily applied to inform school accountability in a fairer and more reliable manner. Finally, the conclusions arising from these analyses are discussed.

4.2. Descriptive analysis

In this section, the data are explored to gain insight into different patterns that can be useful towards the implementation of the more complex statistical models to follow later in this chapter. Table 4.1 summarises the results obtained in the 2004 SIMCE tests of Mathematics and Spanish Language by pupils in Year 8, grouped by gender and the institutional type of the school they attend.

Table 4.1: Summary of descriptive statistics for the 2004 SIMCE test, by school type and gender

		Spanish Language test				Mathematics test			
School type†	**Gender‡**	**Mean**	**Std. Dev**	**Min.**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
	Female	254.01	47.32	97.51	391.67	245.84	45.7	122.56	406.09
	Male	247.3	50.53	98.09	391.67	257.34	48.52	119.47	406.09
State-funded	**Total**	250.83	48.98	97.51	391.67	251.29	47.4	119.47	406.09
	Female	267.57	46.65	101.12	391.67	258.46	45.61	120.42	406.09
Subsidised	**Male**	256.35	49.35	92.86	391.67	266.69	47.37	116	406.09
independent	**Total**	262.16	48.3	92.86	391.67	262.42	46.65	116	406.09
	Female	309.38	38.48	134.03	391.67	309.76	41.81	130.02	406.09
	Male	297.15	43.53	104.66	391.67	317.09	41.69	123.96	406.09
Independent	**Total**	303.15	41.58	104.66	391.67	313.5	41.91	123.96	406.09
	Female	264.97	48.47	97.51	391.67	256.95	48.13	120.42	406.09
	Male	256.03	51.08	92.86	391.67	267.03	49.91	116	406.09
Overall	**Total**	**260.67**	**49.94**	**92.86**	**391.67**	**261.79**	**49.25**	**116**	**406.09**

† Differences in Mathematics and Spanish scores between types of schools are all significant at the 0.001 level.
‡ Differences in Mathematics and Spanish scores between male and female pupils are all significant at the 0.001 level.

In Table 4.1, it is observed that the Spanish Language scores have a slightly higher variation (49.94) than scores in Mathematics (49.25). This is also confirmed by a coefficient of variation of 0.192 (49.94/260.67) in Spanish against the Mathematics tests with a coefficient of variation of 0.188 (49.25/261.79). Examining the totals by gender, it is appreciated that female pupils tend to perform better in Spanish Language than male pupils, who in contrast tend to perform better in Mathematics. Moreover, male pupils' performance in Spanish Language tends to have more variation than female pupils in Spanish and Mathematics and their own performance in Mathematics. Although the Mathematics and Spanish tests are not directly comparable, because they entail diverse score estimation procedures, the comparisons here are only done in terms of variation from the mean, without assuming one-to-one comparability of tests.

Examining the means by school type, it is observed that the lowest means in Spanish Language and Mathematics are recorded for those pupils attending State-funded schools (250.83). Pupils in subsidised independent schools have a higher mean (262.16) than pupils in State-funded schools, with similar levels of variation for both groups. In contrast, pupils in Independent schools tend to outperform pupils in other types of schools by far, with a mean of 303.15 and a lower level of variation.

In the table below, descriptive statistics are reported for the SIMCE tests of Mathematics and Spanish Language sat by pupils in Year 10 in 2006.

Table 4.2: Summary of descriptive statistics for the 2006 SIMCE test, by school type and gender

		Spanish Language test				Mathematics test			
School type†	**Gender‡**	**Mean**	**St. Dev.**	**Min.**	**Max.**	**Mean**	**St. Dev.**	**Min.**	**Max.**
State-funded	**Female**	249.54	49.63	120.43	398.16	235.74	60.65	93.96	426.58
	Male	243.97	50.62	120.75	398.16	248.18	63.25	93.52	426.58
	Total	246.90	50.18	120.43	398.16	241.63	62.21	93.52	426.58
Subsidised independent	**Female**	265.28	49.05	120.71	453.87	255.98	60.15	93.79	426.58
	Male	255.31	49.24	120.43	398.16	264.53	62.07	93.46	426.58
	Total	260.47	49.39	120.43	453.87	260.10	61.23	93.46	426.58
Independent	**Female**	312.22	42.03	120.71	398.16	324.08	47.63	94.96	426.58
	Male	302.71	45.10	120.79	398.16	332.75	51.09	93.65	426.58
	Total	307.37	43.88	120.71	398.16	328.49	49.61	93.65	426.58
Overall	**Female**	262.15	51.37	120.43	453.87	252.53	63.59	93.79	426.58
	Male	254.62	51.82	120.43	398.16	263.49	65.57	93.46	426.58
	Total	**258.53**	**51.72**	**120.43**	**453.87**	**257.80**	**64.78**	**93.46**	**426.58**

† Differences in Mathematics and Spanish scores between types of schools are all significant at the 0.001 level.
‡ Differences in Mathematics and Spanish scores between male and female pupils are all significant at the 0.001 level.

The patterns observed in the scores achieved in 2006 resemble those recorded in 2004. Nevertheless, slightly more variability with respect to 2004 is recorded in the Spanish Language test with a coefficient of variation of 0.2 (51.72/258.53), while in Mathematics, the increase in variability is more noticeable with respect to the previous occasion with a coefficient of variation of 0.251 (64.78/257.8).

When examining the patterns in the data according to the gender of pupils, it is observed that female pupils tend to perform better in Spanish Language and male pupils tend to perform better in Mathematics (as it also occurred in the 2004 SIMCE tests). This tendency is observed across all school types.

Consistent with the results obtained in 2004, the scores in 2006 according to school type follow the same patterns. Pupils in independent schools greatly outperform pupils in subsidised independent schools and pupils in State-funded schools, whereas pupils in subsidised independent schools record fairly higher means than pupils in State-funded schools, with similar levels of variation. In Figure 4.1 below, the distribution of the SIMCE tests scores is summarised and briefly described.

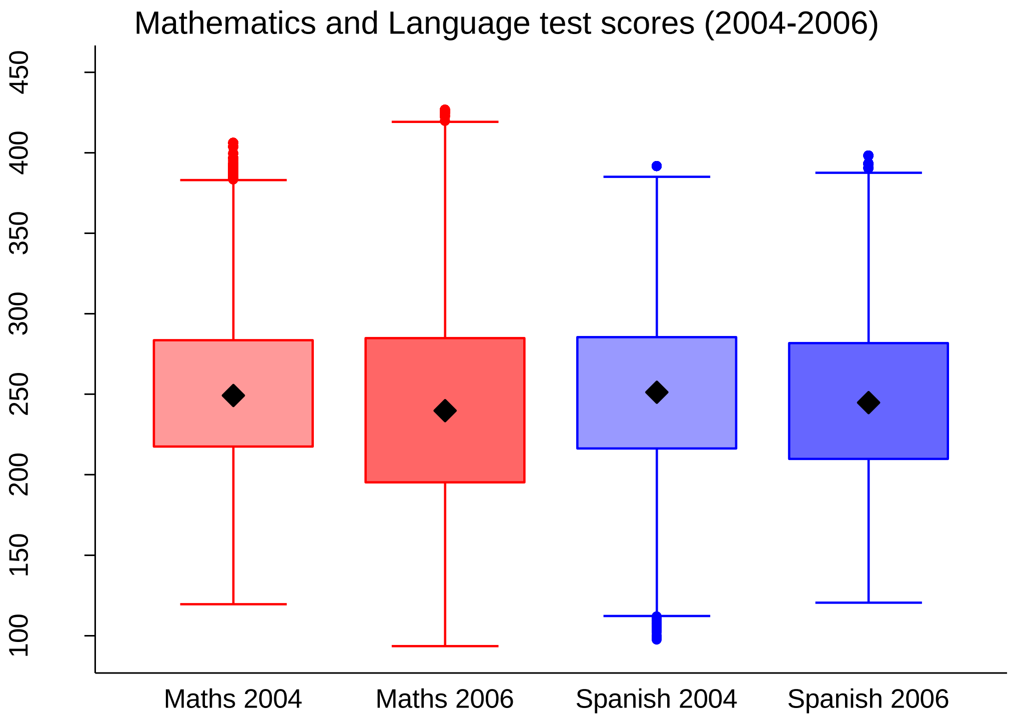


Figure 4.1: Distribution of the Mathematics and Spanish tests scores, 2004-2006

Figure 4.1 above depicts four box plots to summarise the distribution of pupils' scores in both occasions of the Mathematics and Spanish Language SIMCE tests. Firstly, it is observed that the distribution of the scores is much more stable between both occasions of the Spanish Language test in comparison to the Mathematics tests, which record a more spread distribution in 2006. The presence of outliers is noted in all tests, except for the Mathematics test in 2006. In the following graphs, the relationship between the scores on both occasions is depicted.

Figure 4.2 is a scatter plot of prior attainment (x axis) and subsequent attainment (y axis) in Mathematics. It is observed that there is a strong relationship between the scores obtained in the two occasions of the Mathematics test. This justifies empirically the inclusion of prior attainment as a predictor of (current) pupils' performance in Mathematics.

Examining the relationship between the two maths scores, some differences become noticeable, based on the type of school which pupils attended in the second occasion. Independent schools clearly record a higher mean than the other two types of schools and they tend to be clustered around the higher portion of the scatter plot; however, the slope of the line of best fit is less steep than in the other school types, which may indicate that pupils in Independent schools tend to make less progress than other pupils. Nevertheless, the line of best fit does not account for the clustering in the data and hence this must be interpreted with caution.

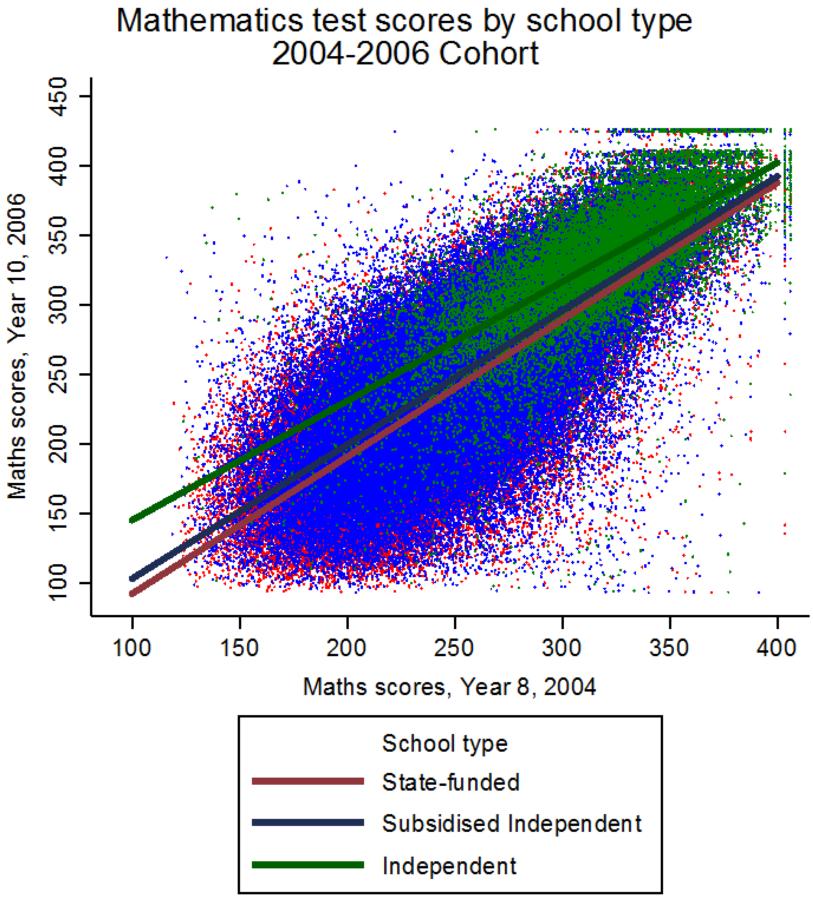


Figure 4.2: Relationship between prior and subsequent attainment in Mathematics, by school type

Comparing the scores of pupils in State-funded schools and subsidised independent schools, it is appreciated that these are much more spread around the whole range of values of the scatter plot, which is due to the higher variability recorded. It is also observed that scores are highly similar, since pupils' scores in both types of schools are almost completely overlaid. When examining pupils' scores in the Spanish tests, very similar patterns are observed. Figure 4.3 depicts this relationship.

As observed in the scatter plot (Figure 4.3), the scores obtained by pupils on both occasions of the Spanish Language tests are, as expected, highly and positively correlated as they were in the case of the Mathematics tests. The visual inspection of the relationship between prior and subsequent attainment in Spanish Language and Mathematics, suggests that schools tend to have different effects on pupils' outcomes.

According to these initial explorations, the implementation of a multilevel model becomes highly relevant, not only because of the nested nature of the data, but also because there seem to be relevant effects at the school level to be analysed.

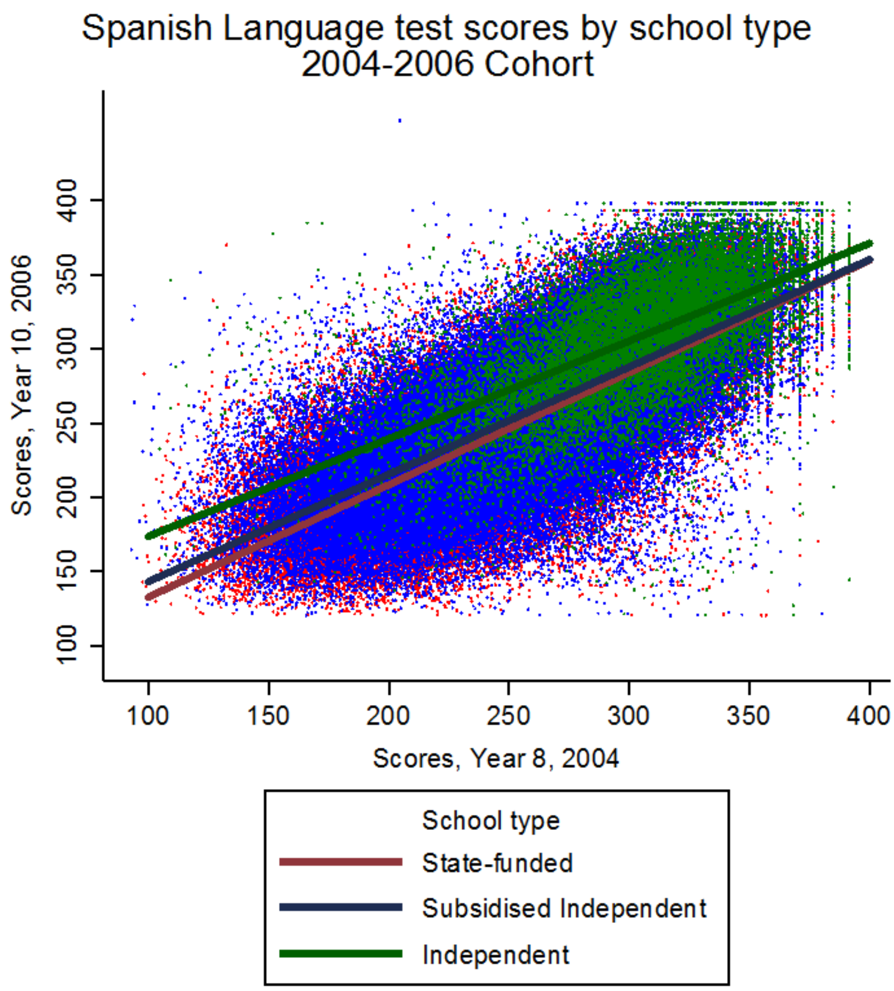


Figure 4.3: Relationship between prior and subsequent attainment in Spanish Language, by school type

Additionally, variation within schools may need to be analysed more thoroughly, since Chilean schools routinely implement streaming measures to boost their performance. This is done by assigning pupils to classes according to their achievement levels. This calls for the necessity of evaluating the specification of an additional level of variation in the multilevel model, i.e. the classroom level.

Likewise, variation between schools may also need to be analysed at a macro level, since there is evidence of regional differences[12]. Although there are known inter and intra-regional inequalities concerning the distribution of wealth and resources, this geographical component of schools' performance has not been greatly explored. In order to address this issue, the addition of a geographically defined macro-level will be assessed, i.e. the local authority level.

[12] ANOVA results indicate that there are highly significant (at the 0.001 level) differences in Mathematics and Spanish scores between the 13 regions of the country.

4.3. Analysing school value-added in Mathematics from a multilevel perspective

4.3.1. Where does the variation in Mathematics scores come from? Variance components models

As outlined in Chapter 3, the first step in this analysis is to specify the most basic multilevel model. This is the empty model of pupils nested within schools, from which more complex value-added models are sequentially built up.

As hinted in the section title, the dependent variable in these models is the standardised Mathematics scores obtained by secondary school pupils in 2006. Table 4.3 shows that the basic 2-level model of pupils nested within schools is significantly improved by the specification of the classroom level, which is indicated by a highly significant chi-square (16,151.7 on 1 degree of freedom) and a lower value for the Akaike Information Criterion (AIC). Likewise, the addition of yet another level, that is the level of local authorities, produced a significant improvement to the overall fit of the model, with a chi-square of 175.5 on 1 degree of freedom and a lower AIC value.

Table 4.3: Summary of the variance components models

Parameter	2-level model	3-level model	4-level model
Fixed part	**Coef. (s.e.)**	**Coef. (s.e.)**	**Coef. (s.e.)**
Intercept	0.018 (0.015)	-0.0003 (0.015)	-0.138 (0.024)
Random part	**Variance**	**Variance**	**Variance**
Pupil	0.57	0.502	0.502
Classroom	--	0.095	0.095
School	0.51	0.47	0.396
Local Authority	--	--	0.069
Log-likelihood	-234542.6	-226466.75	-226379
Deviance	469085.2	452933.5	452758
AIC	469091.2	452941.5	452768
Number of parameters	3	4	5
Chi-square (X^2)	--	16151.7	175.5
Degrees of freedom	--	1	1
N	202,094	202,094	202,094

The addition of an intermediate level, i.e. the classroom level, has reduced the variance of both adjacent levels, i.e. the pupil level (from 0.57 to 0.502) and school level (from 0.51 to 0.47). This finding is consistent with what Martinez (2012) found for US school data, where the addition of the classroom level reduced the variance of both the pupil and the school level. More generally, this is also consistent with Tranmer and Steel (2001), who found, using the UK census data, that the variance proceeding from an intermediate level was redistributed along the two adjacent levels when ignored. Meanwhile, the addition of the local authority level has further reduced the between-school variance (from 0.47 to 0.396). More importantly, it is clear from Table 2 that a 2-level model significantly overestimates the overall school effects.

In Table 4.4, the total variance of the empty models for attainment in Mathematics is decomposed into the different sources of variation, i.e. pupils, classrooms, secondary schools and local authorities.

Table 4.4: Variance partition coefficients for the 2, 3 and 4-level empty models

Level	2-level model	3-level model	4-level model
Pupil	52.78%	47.10%	47.30%
Classroom	--	8.90%	8.90%
Secondary School	47.22%	44.00%	37.30%
Local Authority	--	--	6.50%
Total	**100%**	**100%**	**100%**

In Table 4.4, according to the Variance Partition Coefficient (VPC) in the 2-level model, it is estimated that 47.22% of the total variance is due to the variance between schools. An alternative interpretation is that the average correlation between two randomly selected pupils within the same school is expected to be 0.47. This result is yet another way of demonstrating how a single-level analysis would be severely misleading with a very high intra-class correlation; thus the implementation of a multilevel model is highly relevant. This result is also consistent with Munoz-Chereau (2013) who found a between-school variation of 47.3% in an empty 2-level model for Mathematics attainment in the same cohort of pupils.

This in itself is a remarkable result, since previous research in the United Kingdom has found that schools only account for between 10% and 20% of the total variation (Leckie, 2009, Leckie et al., 2010, Rasbash et al., 2010). This is a sign of the massive differences that can be found in a highly unfair education system, such as the one in Chile.

On another front, the variance partition coefficients for the 3-level model indicate that 44% of the total variation is due exclusively to the between-school variation, whereas the variation exclusively explained by the class level is 8.9%. Analysing the combined VPC (i.e. classrooms and schools), it is appreciated that the class level explains nearly 53% of the total variation. Another way of interpreting this coefficient is that two randomly selected pupils from the same class (who consequently belong to the same school) are expected to have an average correlation of 0.53. This results in an overestimation of the between-school variation by approximately 7% [i.e. (47.22%-44%)/47.22%*100] and approximately an 11% [i.e. (52.78%-47.1%)/52.78%*100] in the case of the between-pupil variation

In turn, in the 4-level model, the variance partition coefficients show that a relevant proportion of the variation that was believed to be due to between-school variability was actually due to the variance between Local Authorities (6.5%); whereas the percentage of the total variation exclusively due to between-school variability is further reduced to a 37.3%. This means that the school effects had been overestimated approximately by 21%, i.e. (47.22%-37.3%)/47.22%*100.

These are remarkable findings, since they reveal that Chilean schools have different levels of performance according to their geographical location. This is most likely due to the Centralism and the highly unequal distribution and concentration of wealth and resources. The addition of the variation between local authorities provides another source of contextual information to this school CVA model, rendering this analysis a geographically and socio-economically contextualised value-added model, as opposed to the traditional CVA models which are only socio-economically contextualised. Furthermore, this is certainly an outstanding difference with respect to what has been found in UK-based research, where Local Authorities only account for approximately 1% of the total variation (Leckie et al., 2010; Rasbash et al., 2010).

These unconditional models are the starting point of the analysis of pupils' and schools' performance; however, these models are not truly value-added models, since pupils' progress is not being measured (Goldstein, 1997). Nevertheless, the specification of the additional levels of classrooms and local authorities has proved to be successful and hence they will be maintained in all subsequent models. In the next section, a measure of prior attainment is specified in order to set up the baseline value-added model.

4.3.2. A baseline model of school value-added in Mathematics

This model is set up by adding prior attainment scores as the only explanatory variable in the fixed part of the 4-level model. The results are presented below in Table 4.5, which shows that the inclusion of prior attainment scores produces a drastic reduction of the variance at the four specified levels. It is also confirmed by the model fit indicators, that this specification is a significant improvement to the overall fit. The coefficient estimated for prior attainment indicates that a one standard deviation change in prior attainment is expected to produce a 0.597 standard deviations change in subsequent attainment. Although this result is relevant to appreciate how highly correlated prior and subsequent attainment are, in terms of school-value added, this model is unsatisfactory as it is shown next.

Table 4.5: Variance components model versus raw value-added model

Parameter	Empty (VC) model	Raw Value-added
Fixed part	**Coef. (s.e.)**	**Coef. (s.e.)**
Intercept	-0.138(0.024)	-0.043(0.012)
Prior attainment	--	0.597(0.002)
Random part	**Variance**	**Variance**
Pupil	0.502	0.291
School	0.396	0.11
Classroom	0.095	0.039
Local Authority	0.069	0.012
Log-likelihood	-226379	-165840
Deviance	452758	331679.8
AIC	452768	331691.8
Number of parameters	5	6
Chi-square (X^2)	--	121078.2
Degrees of freedom	--	1
N	202,094	197,227

Inspecting the school-level residuals estimated from this raw value-added model, it is observed that the raw school value-added estimates seem to be highly correlated with the raw school average attainment scores. This is depicted in Figure 4.4, which is a scatter plot of the estimated school value-added scores against the prior attainment scores averaged at the school level. As other authors have shown (Ferrão & Goldstein, 2009; Goldstein et al., 2007), this is indeed the result of a misspecification bias, which calls for the necessity of including variables to assess underlying differential school effects (Foley & Goldstein, 2012).

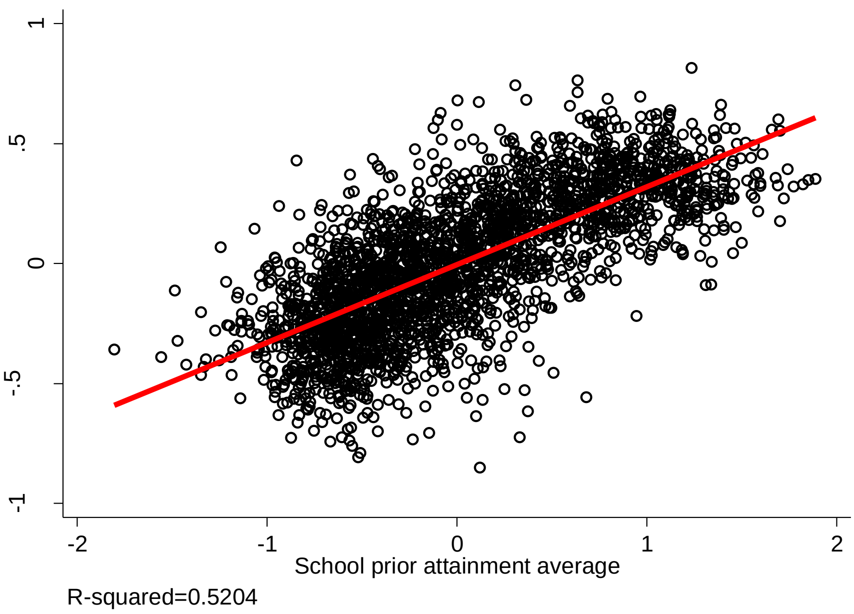


Figure 4.4: Relationship between schools' prior attainment averages and raw value-added estimates

Figure 4.4 shows that the correlation between school prior attainment average scores and their respective raw value-added estimates is clearly not negligible, as there is an obvious pattern where schools serving pupils with higher prior attainment tend to have also higher value-added estimates. This correlation can be thought of as a consequence of positive school-level peer effects and/or selection bias. School peer effects arise when pupils are benefited from a favourable school environment in which pupils perform well, and hence influence other pupils to also perform well. On the other hand, selection bias arises because high-attainers attend high-attending schools. Both biases can be controlled by several specifications. As mentioned in chapter 3, school selectivity is controlled (at least partially) via pupils' prior attainment and year repetition. School average prior attainment has been used in previous research to control for peer effects; however, this is not done in this chapter, as it is shown that controlling for school and pupil socio-economic characteristics reduces this correlation considerably, which renders it not problematic. Additionally, Goldstein et al. (2007) have shown that this misspecification problem can be remedied by specifying random coefficients for prior attainment at the higher levels as well as by adding other contextual (school-level) explanatory variables, as it will be seen later in this chapter.

4.3.3. A contextualised value-added model for progress in Mathematics

Adding pupil-level explanatory variables to the fixed part of the CVA model

As described in the analysis strategy of chapter 3, after specifying the raw value-added model, the explanatory variables at the level of pupils are to be included in the model as the first step to allow for the value-added to have a contextualised form. Only the pupil-level variables are added first to control for compositional effects in the estimated school value-added measures. The variables that were added in this step are: gender, average monthly household income and year repetition.

From the addition of these variables, it is learned that the effect of prior attainment is only slightly reduced from 0.597 to 0.589; however, the significance of the interaction effects renders this main effect less relevant to interpret. As it will be appreciated in subsequent models, this effect is always the largest and only varies slightly. The addition of these individual-level variables significantly improves the overall fit of the model, with a reduction of the deviance of 28,777.65 on 5 degrees of freedom with respect to the raw value-added model presented previously. This model is presented in Table 4.6.

Given that significant effects were found for interactions involving all the explanatory variables, the main effects of the fixed part of the model are not to be interpreted in their own right, but in combination with the variables where the interactions have been found to be significant. For instance, the main effect of gender is significant although it varies by year repetition, income and prior attainment. Likewise, the main effect of year repetition is also significant but moderated by gender and prior attainment.

Regarding the interaction effects, it is observed that the effect of prior attainment varies significantly according to gender, being held back and household income. In the case of gender, it can be appreciated that male pupils make slightly less progress than female pupils, or in other words, the relationship between prior and subsequent attainment in Mathematics has a steeper slope in the case of female pupils. As it will be seen later on in this chapter, this relationship holds even when controlling for all the variables in the full model and is depicted for easier interpretation in Figure 4.7, where it can be observed that male pupils appear to be more advantaged than female pupils, but the gap decreases as prior attainment increases.

Table 4.6: Fixed and random effects for the CVA model with pupil-level explanatory variables only

Fixed Part	**Coef.**	**S.E.**	**z**	**P>z**	**95% Conf. Interval**	
Main effects†						
Intercept	-0.072	0.011	-6.545	0.000	-0.094	-0.050
Prior attainment	0.589	0.003	196.333	0.000	0.583	0.595
Male	0.074	0.004	18.500	0.000	0.066	0.082
Low-mid income	0.027	0.004	6.750	0.000	0.019	0.035
Up-mid income	0.05	0.007	7.143	0.000	0.036	0.064
High income	0.079	0.01	7.900	0.000	0.059	0.099
Held back	-0.269	0.009	-29.889	0.000	-0.287	-0.251
Interaction effects†						
Prior attainment & Male	-0.012	0.003	-4.000	0.000	-0.018	-0.006
Prior attainment & Low-mid income	0.015	0.003	5.000	0.000	0.009	0.021
Prior attainment & Up-mid income	0.025	0.005	5.000	0.000	0.015	0.035
Prior attainment & High income	0.026	0.006	4.333	0.000	0.014	0.038
Prior attainment & Held back	-0.08	0.005	-16.000	0.000	-0.090	-0.070
Male & Low-mid income	-0.013	0.006	-2.167	0.030	-0.025	-0.001
Male & Up-mid income	-0.019	0.009	-2.111	0.035	-0.037	-0.001
Male & High income	0.001	0.011	0.091	0.928	-0.021	0.023
Male & Held back	0.078	0.009	8.667	0.000	0.060	0.096
Low-mid income & Held back	0.013	0.011	1.182	0.237	-0.009	0.035
Up-mid income & Held back	0.017	0.019	0.895	0.371	-0.020	0.054
High income & Held back	0.018	0.024	0.750	0.453	-0.029	0.065
Random part	**Variance**	**S.E.**			**95% Coverage int.**	
Level 4: Local Authority	0.009	0.002	--	--	0.005	0.014
Level 3: School	0.099	0.004	--	--	0.091	0.106
Level 2: Classroom	0.037	0.001	--	--	0.035	0.039
Level 1: Pupil	0.286	0.001	--	--	0.283	0.287
Log-likelihood	-151253.37					
Deviance	302506.73					
AIC	302552.73					
Number of parameters	23					
Chi-square (X^2)	29173.069					
Degrees of freedom	17					
N	181,867					

† Reference categories: Female; Low income and Not held back.

In the case of year repetition, being held back diminishes the effect of prior attainment and thus pupils who have been held back in primary school are expected to make less progress than pupils who have not been held back. Gender also significantly interacts with year repetition. When comparing males and females who have been held back in one or more years, it is observed that boys do significantly better than girls, and hence the negative effect of being held back is greater for girls.

With respect to the interaction terms for the effects of household income and gender, it is observed that the differences between boys and girls significantly vary by the level of income of the household. Male pupils in lower-middle and upper-middle income households are expected to make significantly less progress than female pupils in the same type of households and male pupils in the lowest-income households. Nevertheless, boys and girls in the highest-income households do not significantly differ from each other; neither do they differ from pupils in the lowest-income households. Finally, this model estimates that the effect of year repetition does not differ significantly across income groups, and hence, it is removed from the final model.

Regarding the random part of the model, it is appreciated that all variances have suffered some reductions with respect to the raw value-added model in Table 4.5. In combination with the reductions recorded in the raw value-added model with respect to the gross school effects model (variance components model in Table 4.5), it can be asserted that a great proportion of the total variance is due to pupil-level variables. From the perspective of the secondary schools, it is observed that most of the initially recorded school effects in the empty model are not actually due to the schools themselves, but to individual characteristics and abilities of the pupils.

In the next section, the effects of the school-level variables are to be estimated, whilst keeping the effects of the significant pupil-level variables and their interactions to avoid confounding by compositional effects. This is done with the purpose of estimating more precise school value-added estimates, where the effects of non-malleable school factors are removed.

Adding school-level explanatory variables to the fixed part of the CVA model

In this step, the school-level variables are added to the fixed part of the CVA model. Since all effects (main and interactions) are assumed to be fixed at this point, this contextualised value-added model constitutes a random intercepts model, which is similar to the approach adopted by the UK Department for Education and Skills until 2011 (Leckie & Goldstein, 2009; Ray, 2006). The fixed part of the model is presented in Table 4.7.

As observed in the previous step when adding the individual-level explanatory variables, several interaction effects were found to be significant, which renders the interpretation of the main effects on their own unimportant. Meanwhile, school-level effects have not been allowed to interact and hence their main effects can be interpreted in their own right.

Table 4.7: Fixed effects of the random intercepts CVA model

Fixed Part	Coef.	S.E.	z	P>z	95% Conf. Interval	
Main effects†						
Intercept	-0.319	0.015	-21.267	0.000	-0.348	-0.290
Prior attainment	0.587	0.003	195.667	0.000	0.581	0.593
Male	0.075	0.004	18.750	0.000	0.067	0.083
Low-mid income	0.019	0.004	4.750	0.000	0.011	0.027
Up-mid income	0.031	0.007	4.429	0.000	0.017	0.045
High income	0.038	0.01	3.800	0.000	0.018	0.058
Held back	-0.267	0.009	-29.667	0.000	-0.285	-0.249
Lower-middle SES	0.113	0.018	6.278	0.000	0.078	0.148
Middle SES	0.319	0.02	15.950	0.000	0.280	0.358
Upper-middle SES	0.543	0.024	22.625	0.000	0.496	0.590
Upper SES	0.718	0.045	15.956	0.000	0.630	0.806
Subsidised independent	0.057	0.015	3.800	0.000	0.028	0.086
Independent	0.029	0.041	0.707	0.479	-0.051	0.109
Interaction effects†						
Prior attainment & Male	-0.012	0.003	-4.000	0.000	-0.018	-0.006
Prior attainment & Low-mid	0.014	0.003	4.667	0.000	0.008	0.020
Prior attainment & Up-mid	0.022	0.005	4.400	0.000	0.012	0.032
Prior attainment & High	0.022	0.006	3.667	0.000	0.010	0.034
Prior attainment & Held back	-0.079	0.005	-15.800	0.000	-0.089	-0.069
Male & Low-mid income	-0.012	0.006	-2.000	0.046	-0.024	-0.0002
Male & Up-mid income	-0.018	0.009	-2.000	0.046	-0.036	-0.0004
Male & High income	0.002	0.011	0.182	0.856	-0.020	0.024
Male & Held back	0.077	0.009	8.556	0.000	0.059	0.095
Low-mid income & Held back	0.012	0.011	1.091	0.275	-0.010	0.034
Up-mid income & Held back	0.013	0.019	0.684	0.494	-0.024	0.050
High income & Held back	0.008	0.024	0.333	0.739	-0.039	0.055

† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.

Regarding the school-level variables, it is appreciated that pupils in schools with a low overall socio-economic status (the reference SES category) are expected to perform significantly worse than all other pupils in higher SES schools. More importantly, it is found in this model that state-funded schools do not differ significantly from independent schools.

Analysing the main effects of the fixed part, it is observed that the effects of the pupil-level variables remain almost unchanged, except for minor differences in some coefficients, while significance is the same.

Table 4.8: Random effects of the random intercepts CVA model

Random part	Variance	S.E.	95% Coverage int.	
Level 4: Local Authority	0.005	0.001	0.003	0.007
Level 3: School	0.054	0.002	0.050	0.058
Level 2: Classroom	0.037	0.001	0.035	0.039
Level 1: Pupil	0.285	0.001	0.283	0.287

Table 4.8 shows that the variance components of the CVA model have been reduced when comparing with the random part of the previous models. This is especially noticeable for the case of the school-level variance, which has been reduced from 0.11 as estimated in the raw value-added model (Table 4.5) to 0.054 with the random intercepts contextualised value-added model.

Table 4.9: Model fit comparison between CVA model with level 1 and level 3 explanatory variables

Parameter	CVA model with level 1 variables	CVA model with level 2 variables
Log-likelihood	-151253.37	-150682.8
Deviance	302506.73	301365.6
AIC	302552.73	301423.6
Number of parameters (k)	23	29
Chi-squared (X^2)	--	1163.993
Degrees of freedom	--	6
N	181,867	181,867

In Table 4.9, it can be seen that the model with school-level variables produced a significant reduction of the deviance of 1,141.13 on 6 degrees of freedom with respect to the previous model that included pupil-level explanatory variables only. The value of the Akaike Information Criterion (AIC) is also consistent with the conclusion that this is a better-fitting model.

Although these results are all relevant contributions for explaining pupils' progress in Mathematics and school value-added, this model is still unrealistic, since it does not account for differences in prior achievement at the higher levels of the structure. As other authors have pointed out (Goldstein et al., 2007; Leckie et al., 2010; Leckie, 2009), the complexity of the network of effects on pupils' performance has not been traditionally taken into account in educational research and government policy, which can be misleading since school effects are usually overestimated in simpler models.

A more realistically complex approach to analyse pupils' performance and school value-added includes specifying random coefficients for those covariates that are reasonably believed to vary randomly across the higher levels. In this chapter, random coefficients for prior attainment and gender of the pupils have been specified at all the higher levels of the structure. In the next section, the details of this procedure are presented.

Does the relationship between gender and prior and subsequent attainment vary across classrooms, schools and local authorities?

In this part of the analysis, prior attainment and gender of pupils were allowed to vary randomly at the higher levels of the model. This was done one at a time, at each level, i.e. classrooms, secondary schools and local authorities. The significance of the specification of each random coefficient was assessed by contrasting the resulting deviance and AIC of each model including the specific random coefficient specification to be tested with the deviance and AIC of the previous model, i.e. the random intercepts CVA model (Tables 4.7, 4.8 and 4.9).

The substantive relevance of this specification lies in its potential for assessing differential school effects for diverse groups of pupils. For instance, the inclusion of a random variation component for pupils' prior attainment at the level of secondary schools allows identifying schools that have shallower (or steeper) than average rates of progress. In the same way, specifying a random coefficient for gender of pupils (more specifically for boys), allows identifying secondary schools that do worse for girls than boys (or vice versa). All this information is potentially useful and can be explored further with qualitative research or school inspection.

The first random coefficient specification is prior attainment in Mathematics. This random coefficient produced significant improvements to the overall fit of the model when specified at each higher level and hence six random effects were added to the CVA model, i.e. the variance of the slope residuals for prior attainment at the levels of classrooms, schools and local authorities, and their corresponding covariance terms associated with the variance terms of each random intercept at the specified levels. The second further random coefficients specification is the addition of the variance components for the gender of pupils at all the higher levels of the structure. This produced a total of nine additional random-effects parameters, i.e. the variance of male pupils at the levels of classrooms, secondary schools and local authorities, along with the covariance between those variance terms and the variance terms for the intercept and prior attainment at all levels. In Table 4.10, for the sake of simplicity, only the random part of the model is presented, since the fixed part remained mostly unchanged with respect to the random intercepts CVA model (Table 4.7), except for negligible differences. Detailed results of the fixed part of this model can be found in Appendix 1.1.

In Table 4.10, neither the variance nor the covariance terms can be interpreted separately at any level, except for the pupil-level, which only has an intercept variance. Having noted that, it is best to give more attention to the correlations estimated for the variance terms at each level. At the level of the local authorities, the correlation between the intercept and the random coefficient for male pupils is -0.532, which indicates that the higher the local authority intercept, the less steep is male pupils' progress. In turn, the correlation between the intercept and the slope for prior attainment is very high (0.842), which indicates that the higher the intercept for a particular local authority, the steeper the slope of prior attainment. The correlation between the random coefficient for male pupils and the slope of prior attainment is -0.576, which indicates that the bigger the difference between males and females in a particular local authority, the less pronounced the slope for prior attainment.

At the secondary school level, the correlation coefficient between the random coefficient for male pupils and the intercept is 0.345, which indicates that the higher the intercept of a particular school, the steeper is male pupils' progress in comparison to female pupils, although this is not a strong relationship. With regard to the correlation between the intercept and the slope for attainment, the relationship is not too strong (0.140); however, it does indicate that schools with high intercepts tend to have slightly steeper slopes for attainment. Finally, the correlation between the random coefficient for male pupils and the slope for prior attainment is rather weak (-0.079), and hence not too meaningful.

Table 4.10: Random effects of the random coefficients CVA model without cross-level interaction effects.

Parameter	Estimate	S.E.	Correlation	95% Coverage int.	
Level 4: Local Authority					
Variance (Intercept)	0.006	0.002	--	0.003	0.009
Covariance (Prior attainment, Intercept)	0.002	0.0004	0.842	0.001	0.002
Variance (Prior attainment)	0.001	0.0002	--	0.0002	0.001
Covariance (Male, Intercept)	-0.0003	0.0004	-0.532	-0.001	0.0004
Covariance (Male, Prior attainment)	-0.0001	0.0001	-0.576	-0.0004	0.0001
Variance (Male)	0.0001	0.0002	--	-0.0003	0.0004
Level 3: School					
Variance (Intercept)	0.054	0.003	--	0.049	0.059
Covariance (Prior attainment, Intercept)	0.002	0.001	0.140	0.0005	0.003
Variance (Prior attainment)	0.003	0.0003	--	0.002	0.004
Covariance (Male, Intercept)	0.003	0.001	0.345	0.001	0.005
Covariance (Male, Prior attainment)	-0.0002	0.0004	-0.079	-0.001	0.001
Variance (Male)	0.001	0.001	--	0.00005	0.003
Level 2: Classroom					
Variance (Intercept)	0.039	0.001	--	0.036	0.041
Covariance (Prior attainment, Intercept)	0.004	0.001	0.303	0.003	0.005
Variance (Prior attainment)	0.004	0.0004	--	0.003	0.005
Covariance (Male, Intercept)	-0.004	0.001	-0.182	-0.006	-0.002
Covariance (Male, Prior attainment)	-0.0001	0.001	-0.021	-0.001	0.001
Variance (Male)	0.011	0.001	--	0.008	0.013
Level 1: Pupil					
Variance (Intercept)	0.278	0.001	--	0.276	0.280

At the level of classrooms, the correlation between the random coefficient for boys and the intercept is -0.182, which is not too strong; nevertheless, it does seem to indicate that the higher the intercept of a given classroom the more similar the progress of male and female pupils. With respect to the relationship between the slope for attainment and the intercept, the correlation is 0.303, which would be indicating that the higher the intercept of a classroom, the steeper the slope for prior attainment. As noted at the level of schools, the correlation between the random coefficients for prior attainment and gender is very weak (-0.021) and therefore negligible.

In Table 4.11 below, it can be observed how this random coefficients CVA model has significantly improved the overall fit in comparison to the random intercepts CVA model.

Table 4.11: Model fit comparison between random intercepts CVA model and random coefficients CVA model

Parameter	RI-CVA model	RC-CVA model
Log-likelihood	-150671.37	-150192.42
Deviance	301342.74	300384.83
AIC	301410.74	300466.83
Number of parameters (k)	29	41
Chi-squared (X^2)	--	957.91
Degrees of freedom (df)	--	12
N	181,867	181,867

Although this model represents a significant improvement, the significance of the variance components for prior attainment and gender makes the existence of significant cross-level interactions effects plausible. The next model is the full random coefficients contextualised value-added model, where cross-level interaction effects are estimated along with all the specifications made up to this point.

4.3.4. An extended random coefficients CVA model: Full model including cross-level interaction effects

When interacting prior attainment with the school level variables, i.e. institutional type and school SES, significant coefficients were estimated. On the contrary, none of the cross-level interaction terms between gender and the school-level variables turned out to be significant, and hence, they were removed from the final model. Detailed results of this sub-step can be inspected in Appendix 1.3. Results from the full final CVA model for Mathematics progress are presented in detail below in Table 4.12.

In Table 4.12, it is observed that the main effects of the fixed part have remained mostly unchanged, while significance does not vary with respect to the previous models. Regarding the interaction effects of the fixed part, these remain unchanged with respect to the previous models, with the exception of the interaction between prior attainment and income, which was removed from the final model, since none of the coefficients were significant. Detailed results can be inspected in Appendix 1.2. This is presumably the effect of specifying the cross-level interaction between prior attainment and school SES, where significant differences were found. This could be said to imply that the schools' socio-economic sorting plays a more important role than the pupils' own socio-economic background.

With regard to the cross-level interaction effect between school SES and pupils' prior attainment, it is found that the effect of prior attainment is more pronounced for pupils in middle and upper-middle SES schools than for their counterparts in low SES schools. Meanwhile, the effect of prior attainment did not turn out to be significantly different between low SES schools, lower-middle SES schools and upper SES schools. This implies, for instance, that a pupil with low prior attainment will be better off at a middle SES or an upper-middle SES school, whereas they would make the same expected progress in Mathematics in either a low SES, a lower-middle SES or an upper SES school.

Table 4.12: Fixed effects of the full CVA model

Main effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Intercept	-0.316	0.015	-20.531	0.000	-0.346	-0.286
Prior attainment	0.582	0.006	96.109	0.000	0.571	0.594
Male	0.064	0.004	17.497	0.000	0.057	0.071
Lower-middle income	0.012	0.003	3.665	0.000	0.006	0.018
Upper-middle income	0.024	0.005	4.547	0.000	0.013	0.034
High income	0.041	0.008	5.354	0.000	0.026	0.056
Held back	-0.258	0.008	-33.457	0.000	-0.273	-0.243
Lower-middle SES	0.113	0.018	6.184	0.000	0.077	0.149
Middle SES	0.330	0.021	15.923	0.000	0.290	0.371
Upper-middle SES	0.549	0.025	22.257	0.000	0.501	0.597
Upper SES	0.746	0.047	15.963	0.000	0.654	0.838
Subsidised independent	0.063	0.015	4.144	0.000	0.033	0.093
Independent	0.027	0.043	0.627	0.531	-0.058	0.112
Interaction effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Prior attainment & Male	-0.011	0.003	-3.651	0.000	-0.017	-0.005
Prior attainment & Held back	-0.067	0.005	-12.965	0.000	-0.077	-0.057
Male & Held back	0.081	0.009	8.996	0.000	0.064	0.099
Prior attainment & Lower-middle SES	0.0003	0.007	0.046	0.963	-0.013	0.014
Prior attainment & Middle SES	0.044	0.008	5.678	0.000	0.029	0.059
Prior attainment & Upper-middle SES	0.049	0.010	5.105	0.000	0.030	0.067
Prior attainment & Upper SES	0.017	0.020	0.834	0.404	-0.023	0.056
Prior attainment & Subs. independent	0.013	0.005	2.514	0.012	0.003	0.024
Prior attainment & Independent	0.021	0.019	1.139	0.255	-0.015	0.058

† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.

In turn, when inspecting the estimated coefficients for the cross-level interaction effect between pupils' prior attainment and school type, it is found that the effect of prior attainment is more pronounced for pupils in subsidised independent schools in comparison to pupils in state-funded schools. This would imply for instance that a pupil with low prior attainment would be better off in a subsidised independent schools, were their parents able to afford any fees (should there be in place in that particular school), i.e. the pupil would progress more in Mathematics. More interestingly, independent schools with the same level of prior attainment of State-funded schools are not expected to differ significantly in subsequent tests; which would imply that independent schools do not significantly add value to their pupils' educational trajectories in Mathematics when comparing them to State-funded schools.

From the analysis of both cross-level interaction effects, it is apparent that the advantage of attending an independent school does not reside in the potential progress that pupils would make, should they be able to afford the fees, neither in the alleged socio-economic advantages of attending a school with pupils in the same social stratum. It seems more likely that independent schools take advantage of their ability to shape their pupil intake. In simpler terms, given that independent schools are allowed to select the most able pupils, their attainment in subsequent tests is high, but pupils do not make more progress than what they would be expected to make, if they attended a State-funded school.

In table 4.13 below, the estimated random effects of the full CVA model are presented. These random effects do not differ meaningfully from the random effects estimated for the model without cross-level interactions (Table 4.10).

Table 4.13: Random effects of the full CVA model

Parameter	Estimate	S.E.	Correlation	95% Coverage int.	
Level 4: Local Authority					
Variance (Intercept)	0.007	0.002	--	0.004	0.011
Covariance (Prior attainment, Intercept)	0.002	0.001	0.896	0.002	0.003
Variance (Prior attainment)	0.001	0.0002	--	0.001	0.001
Covariance (Male, Intercept)	-0.0004	0.0004	-0.464	-0.001	0.0005
Covariance (Male, Prior attainment)	-0.0001	0.0002	-0.358	-0.0004	0.0002
Variance (Male)	0.0001	0.0002	--	-0.0003	0.0005
Level 3: School					
Variance (Intercept)	0.053	0.003	--	0.048	0.058
Covariance (Prior attainment, Intercept)	0.001	0.001	0.111	0.00002	0.003
Variance (Prior attainment)	0.002	0.0003	--	0.002	0.003
Covariance (Male, Intercept)	0.003	0.001	0.354	0.001	0.005
Covariance (Male, Prior attainment)	-0.0001	0.0003	-0.039	-0.001	0.001
Variance (Male)	0.001	0.001	--	0.0001	0.003
Level 2: Classroom					
Variance (Intercept)	0.039	0.001	--	0.037	0.042
Covariance (Prior attainment, Intercept)	0.004	0.001	0.315	0.003	0.005
Variance (Prior attainment)	0.004	0.0004	--	0.003	0.005
Covariance (Male, Intercept)	-0.004	0.001	-0.181	-0.006	-0.002
Covariance (Male, Prior attainment)	-0.0001	0.001	-0.019	-0.001	0.001
Variance (Male)	0.011	0.001	--	0.008	0.013
Level 1: Pupil					
Variance (Intercept)	0.278	0.001	--	0.276	0.280

Highlighting key aspects of the estimated random effects for the full model, it can be asserted that at the level of schools, male pupils are expected to make more progress than female pupils when the school they attend has a higher intercept. In other words, high-achieving secondary schools are more effective for boys than for girls. Nevertheless, the correlation recorded for this relationship is only moderate. On the other hand, the effect of prior attainment is greater when the intercept of a particular secondary schools is higher, or in simpler terms, high-achieving secondary schools tend to make their pupils progress slightly more. Nevertheless, this is not a strong pattern whatsoever; it is indeed weaker than the relationship between being male and the school intercept. Finally, the effects of prior attainment and gender seem to be rather weakly correlated, indicating that high achievers tend to make more progress regardless of their gender and that male pupils tend to do better in Mathematics, regardless of their prior level of attainment.

Using the baseline raw value-added model (Table 4.5), it is possible to estimate a measure of the explanatory power for progress in Mathematics. It is found that this random coefficients CVA model accounts for 50.91%[13] of the variation in Mathematics progress at the school level, while only a 4.47%[14] of the variance at the pupil level[15]. One very likely possibility is that there are other relevant effects related to the characteristics of the pupils, their family backgrounds and their schools, which have not been controlled for yet in this model. In chapter 6, the effects of a latent variable representing cultural capital will be analysed in more detail, as a way to explore pupils' heterogeneity further and to search for alternative internal school accountability measures.

With respect to the empty model (Table 4.5), the full CVA model has a higher proportion of variance explained at all levels. In the case of the pupil-level variance, this has been explained in a 44.62%[16]. The variance of the classroom has been explained in a 61.05%[17]. The variance between secondary schools has been accounted for in an 86.36%[18]. Lastly, the variance between local authorities has been explained in a 92.75%[19]. The difference between the variance explained at all levels with respect to the empty model and the raw value-added

[13] This is calculated in the following way: [(0.11-0.054)/0.11]*100. Given the presence of the random coefficients, the total school-level variance depends also on prior attainment and gender of the pupils. This is why the school-variance of the full model has not been used for this procedure; the value used corresponds to the school variance of the random intercepts model of Table 4.8.

[14] This is calculated in the following way: [(0.291-0.278)/0.291]*100.

[15] The variance accounted for at the local authority level is 58.33%, while the variance at the classroom level has only been accounted in a 5.13%. This is by no means a concern, since the aim of this research is set at the school and pupil level and hence, these random effects should only be checked for meeting model assumptions.

[16] This indicator is calculated as follows: [(0.502-0.278)/0.502]*100

[17] This indicator is calculated as follows: [(0.095-0.037)/0.095]*100

[18] This indicator is calculated as follows: [(0.396-0.054)/0.396]*100

[19] This indicator is calculated as follows: [(0.069-0.005)/0.069]*100

model seems to imply that a great deal of the variation in Mathematics scores is highly dependent upon prior attainment. Once prior attainment is accounted for (raw value-added model), the remaining variation is deemed to be low, especially at the levels of local authorities and classrooms.

Not surprisingly, Table 4.14 shows that the full CVA model including cross-level interaction effects is indeed a better-fitting model than the previous CVA model presented. The CVA model without cross-level interactions is only displayed in relation to its random effect in Table 4.10 (fixed-effects parameters are available in Appendix 1.1), while the CVA model with cross-level interaction effects is shown in Table 4.12.

Table 4.14: Model fit comparison between CVA model with and without cross-level interaction effects

Parameter	CVA without cross-level interactions	CVA with cross-level interactions
Log-likelihood	-150192.42	-150151.37
Deviance	300384.83	300302.74
AIC	300466.83	300384.74
Number of parameters (k)	41	41
N	181,867	181,867

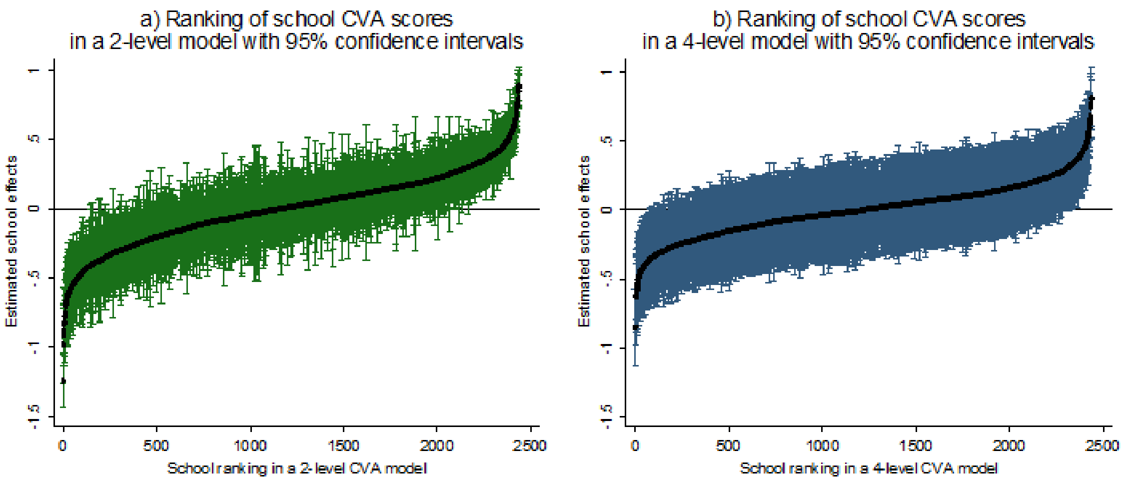
The likelihood ratio test cannot be computed from this model fit comparison since the number of parameters is the same; nevertheless, both models are indeed different insofar as the cross-level interaction effects of the full model override the estimated coefficients for two pupil-level interaction effects that were significant previously. The values of the log-likelihood and deviance are displayed for reference only, while the truly relevant statistic to focus on is the AIC, which shows that the full CVA model does improve the overall fit when compared to the CVA model without the cross-level interaction effects.

In the next section, some of the practical implications of the estimated full CVA model will be analysed.

4.3.5. A practical application of the extended CVA model to inform school performance

After arriving to a satisfactory CVA model, where school effects are controlled for a number of explanatory variables and identified or isolated from random effects at other levels, the analysis turns to examine school value-added more thoroughly in terms of its practical usefulness. The first step is to check whether residuals at all levels meet the assumption of normality, which can be consulted in Appendix 5. Residuals at the school level (i.e. school effects) are considered to be approximately normally distributed.

Having ascertained that assumptions are met after the specification of the extended 4-level CVA model, a reasonable question arises regarding the necessity of the specified complexity. Traditional 2-level CVA models have been around for decades and have been used extensively with the purpose of informing school accountability, and hence, their usefulness is proved. This begs the question of how this extended CVA model contributes to making fairer and/or more precise school comparisons. To address this question, a 2-level CVA model was fitted with the same specifications of the full 4-level CVA model, except for the specification of the random effects at the levels of the classrooms and the local authorities. Both models do not differ greatly in terms of the fixed-effects parameters; however, the random parts have important differences, which are obviously related to the diverse specifications of the number of levels as also seen in Tables 4.3 and 4.4. A more detailed comparison of the coefficients of both models can be found in Appendix 1.4.



Note: The reference line at 0 in the Y-axis of both figures represents the national average. A school is different from the mean when its 95% CI does not overlap with this reference line

Figure 4.5: Comparison of school performance rankings based on CVA scores derived from a 2-level model and a 4-level model

As pointed out in Chapter 3, residual point estimates and their corresponding standard errors are useful in multilevel models to make comparisons between the effect of the higher level units and the overall average. School residuals and their corresponding confidence intervals are estimated by using the shrunken residuals formulae described by Goldstein (Goldstein, 2011), extended to the 4-level model case. Since the residual point estimates at the school level are said to represent the amount of value-added that schools have in comparison to others and considering that the model assumes that the overall average is equal to zero, when estimating 95% confidence intervals with their estimated standard errors, one is also analysing whether the effect of a particular school on its pupils is significantly above the overall mean or national average, in the case of this research. In Figure 4.5 above, a comparison is depicted between the school performance measures derived from the traditional 2-level CVA model and the extended 4-level CVA model.

In Figure 4.5, the x axis represents the rank of the schools according to their residual point estimates, plotted in the y axis. Furthermore, the bars represent 95% confidence intervals and the line across the y axis at 0 represents the national average progress in Mathematics. It can be appreciated that the estimated CVA scores for the Chilean schools analysed in this chapter vary greatly according to the model that is fitted. In the case of the traditional CVA model (Figure 4.5a, left), it is observed that there is a larger number of schools performing significantly below or above the national average. In contrast, with the extended CVA model (Figure 4.5b, right), the number of average-performing schools is remarkably larger. Movements in the rankings are also large when using the 4-level model compared with the 2-level model. From a total of 2,429 secondary schools, 24% of them moved at least 100 places upwards in the ranking table derived from the 4-level model with respect to the ranking table from the 2-level model. These differences are somewhat difficult to visualise in these populated plots, and hence they are explored in more depth in Table 4.15.

4.15: Comparison of school classifications in the traditional 2-level CVA model and the extended 4-level CVA model

	Extended CVA model				
Traditional CVA model	**Below average**	**Average**	**Above average**	**Total**	**Percentage‡**
Below Average	237	323	0	560	22.92%
Average	4	1,247	1	1,252	51.25%
Above average	0	389	242	631	25.83%
Total	241	1,959	243	2,443	100%
Percentage†	9.86%	80.19%	9.95%	100%	

Note: The diagonal shows the agreement between the models. 1,726 (70.65%) schools remain in the same category. The correlation between school effects estimates from both models is 0.964.
† Within extended CVA model classifications
‡ Within traditional CVA model classifications

Based on whether the confidence interval of a school CVA estimate overlaps or not with the national average, a simple 3-level classification can be derived: 1) schools below the national average; 2) average schools; and 3) schools above the national average. In Table 4.15, the classifications derived from the traditional CVA model (2-level model) and the extended CVA models (4-level model) are presented for comparison.

With the traditional model, 51.25% of schools are classified as average compared to 80.19% classified as average when using the extended CVA model. In the 2-level model, 25.83% of schools are found to outperform the national average, while 22.92% perform below the national average. In a 4-level model, these results are 9.95% and 9.86%, respectively. On another front, both models agree on 70.65% of the school classifications as noted in Table 4.15.

This shows that there are considerable differences in the way in which schools are classified between the traditional CVA model and the extended CVA model. One could argue that the traditional model misclassifies a large proportion of schools at both ends of the distribution, and hence, it is misleading, but, naturally, there are two sides to this argument. On the one hand, at the lower end of the distribution, many schools are unfairly classified as below average when using a 2-level model, whereas they might be considered as average when using a 4-level model. On the other hand, at the higher end of the distribution, many schools that could be considered as above average in a 2-level model, are classified as average in a 4-level model.

As mentioned in Chapter 2, the main concern of a school accountability system should be fairness to all schools (See for example: Fitz-Gibbon, 1997; OECD, 2008), which implies that schools should ideally be assessed based only on the circumstances over which they have control. Since there is strong evidence of geographical differences (as seen in Table 4.3), holding everything else in the models constant, one should be inclined to regard the classifications derived from the extended CVA model as fairer than the classifications derived from the traditional CVA model. In sum, these results show that the way in which a CVA model is specified is a sensitive matter, which can have significant policy implications when reporting back to diverse stakeholders.

A closer examination of differential school effects

Previously, it has been shown that the school effects estimated from a raw value-added model (section 4.3.2), i.e. controlling for prior attainment only, are severely biased. It was observed that schools serving the initially highest achievers were the ones that obtained the highest CVA estimates (Figure 4.4). In Figure 4.6 below, it is shown how the extended 4-level CVA model effectively reduces this bias, providing a more reliable set of indicators of schools' value-added and pupils' progress in Mathematics.

Figure 4.6 depicts the relationship between the estimated school residuals from the full 4-level CVA model and the school averages in the previous occasion of the Mathematics tests. It is appreciated that even though there is a somewhat upwards trend, the average prior attainment of a school is not a strong predictor of the school effects estimated in the CVA model. This picture differs considerably from the one presented previously (Figure 4.4.), where raw value-added estimates were indeed more strongly correlated with the school average prior attainment in Mathematics.

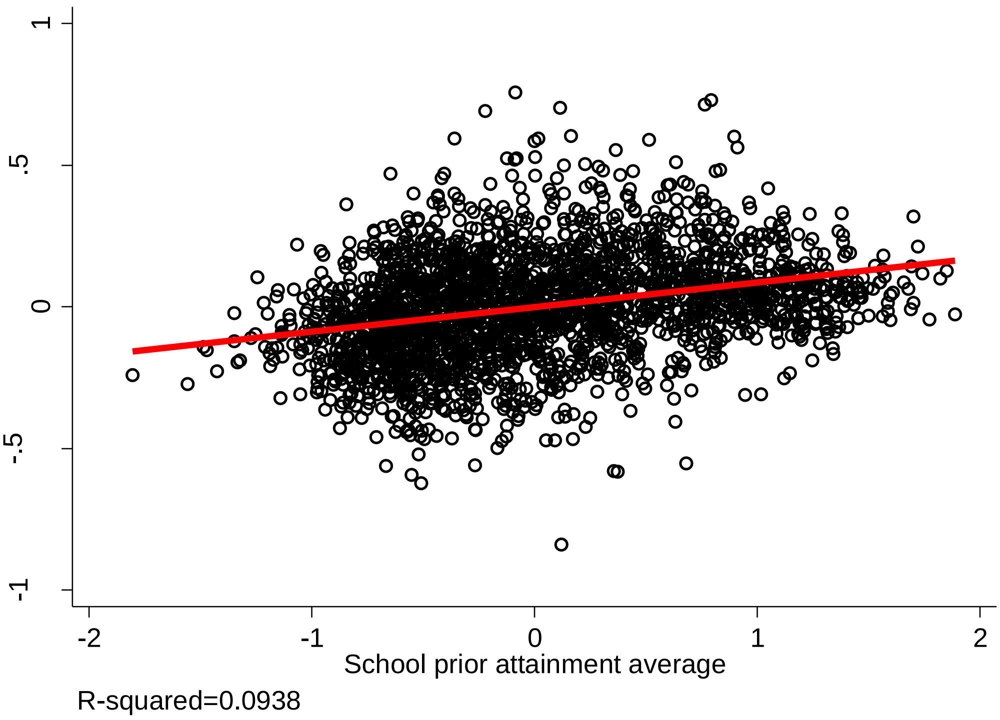


Figure 4.6: Relationship between school prior attainment averages and contextualised value-added estimates

Despite the fact that the correlation shown in Figure 4.6 is not a huge concern in terms of bias of the estimated school effects towards those with higher prior attainment, there still remain other sources of differential school effects across schools, which were specified in the full model to explore more thoroughly. Previously in Chapter 2, it was mentioned that contextualised value-added models can be useful to explore differential school effects. This is achieved by specifying random coefficients for pupil-level explanatory variables at the level of secondary schools as described in step 5 of the analytical strategy (section 3.7.2) of Chapter 3.

Figure 4.7 depicts the relationship between prior and subsequent attainment after controlling for all the variables in the 4-level random coefficients CVA model. It is appreciated here, that, *caeteris paribus*, the differences in scores of male and female pupils are very small and they tend to decrease as prior attainment increases. In other words, after controlling for the fixed effects of income, year repetition, school SES and institutional type, as well as the random effects of classrooms, schools and local authorities, boys and girls have quite similar levels of progress in Mathematics. It can be appreciated that the 95% confidence intervals only overlap at about 1.5-2 standard deviations above the mean of prior attainment. This is crucial information for educational policy, since it is indicative that the gender gap is greater for the less able pupils, and therefore, remedial actions should be pursued in this group of pupils.

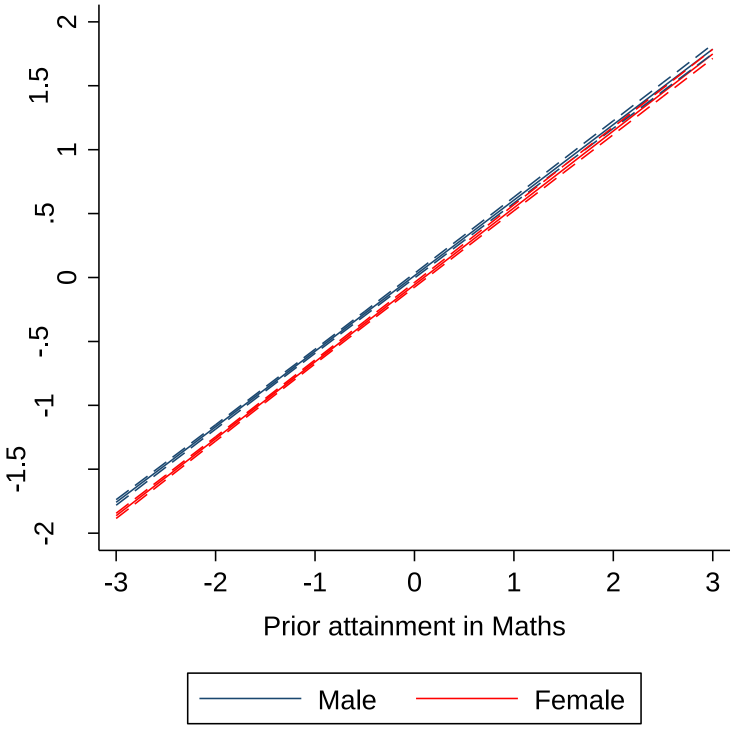


Figure 4.7: Expected progress in Mathematics by gender

Figure 4.8 depicts the relationship between prior attainment and the scores as predicted in the full CVA model grouping pupils according to the average socio-economic status of the school they attend. The predicted scores control for all other variables in the model through the predicted intercept, and hence reliable comparisons can be made. Confidence intervals are omitted for simplicity and readability. The significance of the coefficients can be inspected in Table 4.12.

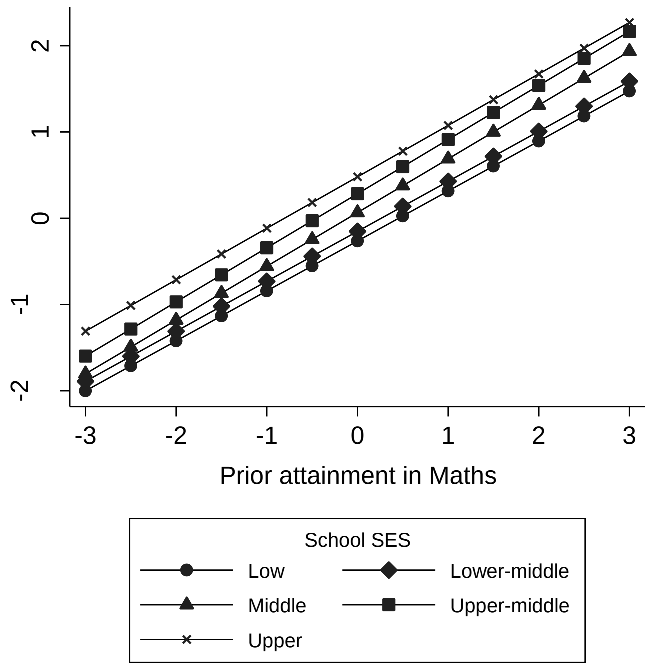


Figure 4.8: Expected progress in Mathematics by school SES

In Figure 4.8, it is observed that after controlling for gender, year repetition, prior attainment, household income, institutional type and all relevant interaction terms, there are diverse rates of progress for pupils depending upon the average socio-economic status of the school to which they attend. It is appreciated that pupils in low and lower-middle SES schools have a slightly shallower progress than pupils in middle, upper-middle and upper-SES schools. It can also be observed that pupils in middle-SES schools have the steepest progress, because even though at the low end of the distribution of prior attainment, these pupils do not differ from pupils in lower-SES schools, they do "catch up" with the higher-SES school pupils at the high end of the distribution of prior attainment. On the other hand, upper-middle SES school pupils have a steady progress in comparison with upper-SES school pupils who start off high and decrease their rate of progress as prior attainment increases.

In Figure 4.9, the depicted relationship between prior and subsequent attainment according to year repetition in primary school already controls for all other variables in the CVA model. A distinct pattern is observed here, where pupils who have been made to repeat at least one year during primary school are expected to make significantly less progress than pupils who have undergone primary school without repeating any year.

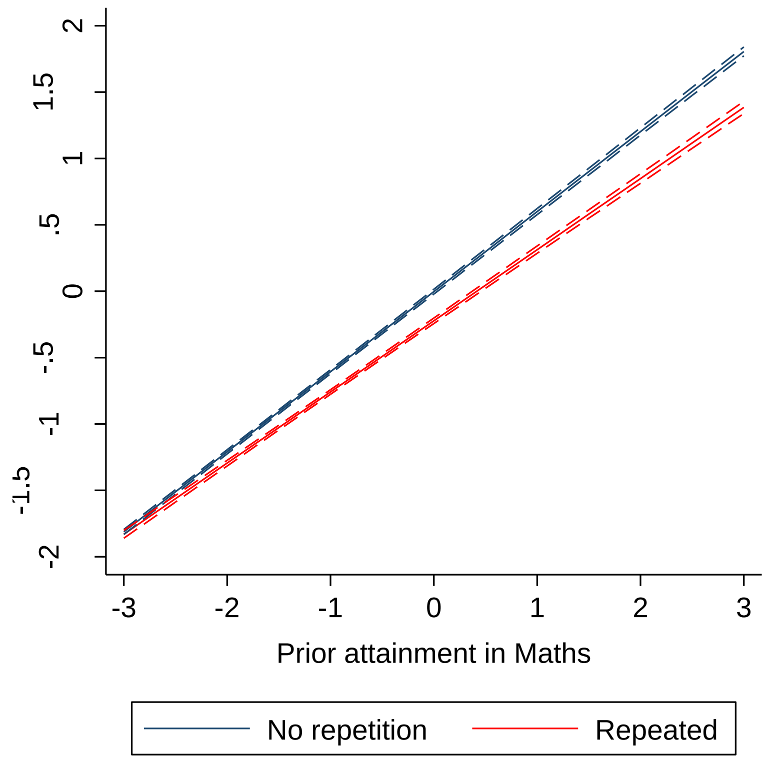


Figure 4.9: Expected progress in Mathematics by pupils' year repetition

In Figure 4.9, the harmful effect of year repetition is apparent. Repetition has no distinguishable effect only at the very extreme low end of the distribution of prior attainment; however, after 2 standard deviations below the mean of prior attainment (still extremely low), the rate of progress of pupils who have been made to repeat a year in primary school is markedly reduced all across the rest of the distribution of prior attainment. This implies that even when pupils have good prior attainment scores in Mathematics, their progress is expected to be below the progress made by pupils who have not repeated any year and have the same level of prior attainment. Furthermore, when inspecting Table 4.12, it can also be appreciated that this effect is even more harmful towards female pupils when interacting year repetition and gender. This certainly constitutes useful information for eventually revising school policy regarding year repetition.

The underlying issue regarding year repetition in primary school is that diverse educational trajectories can have an important impact on secondary schooling. This issue is introduced in the next section and analysed more deeply in Chapter 5.

4.3.6. Carry-over random effects on Mathematics unaccounted for by the 4-level CVA model

As hinted in the analytical strategy outlined in the methodology chapter, the most obvious further extension of this model is the one that specifies the random effects of attending different primary schools before attending secondary schools. One of the problems and/or limitations of the 4-level CVA model presented previously is that it assumes that all pupils have only attended one school during their entire educational trajectory. This can be regarded as highly unrealistic (Fielding et al., 2006; Goldstein et al., 2007).

A suitable approach to account for these carry-over effects from primary schools into secondary school performance is to extend the CVA model further into a cross-classified model, where primary schools are another higher level in the data structure, and thus their effects can be estimated. However, this structure is not fully-hierarchical anymore, which makes this extension not straightforward to implement or readily accessible for a non-statistical audience. Table 4.15 shows the results from the estimated empty 4-level model and the 5-level cross-classified model.

In Table 4.16, it is seen that the variance at the level of primary schools is considerably lower than all other variance estimates. However, the goodness of fit indices show that this further specification of the CVA model is indeed an improvement to the overall fit of the model. The difference between the DIC statistics of both models (433,657.1-431,852.3=1,804.8) renders the specification of the primary school level highly significant. Furthermore, it is observed that the addition of primary schools as a random effect has slightly reduced the variances at the level of pupils, classrooms and schools, which shows that some bias was still present in the estimated random effects of the 4-level CVA model. This model will be further implemented and discussed in detail in Chapter 5 of this thesis.

Table 4.16: Comparison between an empty 4-level fully-hierarchical model and an empty 5-level cross-classified model for Mathematics

Parameter†	Empty 4-level model	Empty 5-level cross-classified model
Fixed part	**Coef. (95% C.I.)**	**Coef. (95% C.I.)**
Intercept	-0.139 (-0.183; -0.093)	-0.16 (-0.202; -0.112)
Random part	**Variance**	**Variance**
Pupil	0.501	0.492
Secondary School	0.396	0.367
Primary School	--	0.015
Classroom	0.095	0.094
Local Authority	0.071	0.071
Deviance ($\overline{\mathbf{D}}$)	427182.9	423362.1
Deviance ($\overline{\boldsymbol{\theta}}$)	420708.8	414871.9
Eff. No of parameters (pD)	6474.16	8490.2
DIC	433657.1	431852.3
N	198,929	198,929

† These parameters were obtained via MCMC estimation with Gibbs sampling, using IGLS estimates as starting values. The chain monitoring length was 20,000, while the burn-in period was 2,000, storing all iterations. Displayed coefficients correspond to the posterior means.

Up to this point, the analyses have only considered progress in Mathematics as the outcome. This begs the question of whether what was found for this subject is also applicable to other subjects. In the next section, progress in Spanish Language is analysed by implementing a contextualised value-added model with the same characteristics of the full CVA model for Mathematics presented in Table 4.12.

4.4. Does progress in other subjects follow the same patterns found in Mathematics? A contextualised value-added model for progress in Spanish Language

The successfully implemented CVA model for analysing progress made by Chilean pupils in Mathematics was specified for the case of attainment in the Spanish Language test, using the same variables on the same group of pupils. Full details are given in Table 4.17.

In Table 4.17, the CVA model for progress in Spanish includes the same fixed effects as in the full CVA model for Mathematics (Table 4.12), and since all scores have been standardised, the effects are comparable in units of standard deviations from the overall mean of each subject (but not in terms of actual achievement). Although this model presents roughly similar patterns in attainment, there are key aspects that are worth pointing out. For instance, the effect of gender is the opposite in the model for progress in Spanish, where male pupils are expected to score lower than female pupils. However, the effect of gender is moderated by year repetition and prior attainment, and hence its main effect is not to be interpreted in its own right. Nevertheless, the effect of repetition on male pupils is the opposite in Language as in Mathematics; boys are more affected in Language when made to repeat than girls. Furthermore, the effect of income seems to be smaller in Spanish progress than in Mathematics, where pupils living in non-low income households are expected to score between 0.061 and 0.579 standard deviations more than pupils in low income households as opposed to Mathematics, where pupils in non-low income households are expected to score between 0.113 and 0.746 standard deviations more than pupils in low-income households (see: Table 4.12).

Table 4.17: Fixed part of the full CVA model for progress in Spanish Language

Main effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Intercept	-0.204	0.011	-17.777	0.000	-0.226	-0.181
Prior attainment	0.664	0.006	115.430	0.000	0.652	0.675
Male	-0.027	0.004	-6.772	0.000	-0.035	-0.019
Lower-middle income	0.032	0.004	8.623	0.000	0.025	0.039
Upper-middle income	0.061	0.006	10.312	0.000	0.050	0.073
High income	0.073	0.009	8.370	0.000	0.056	0.090
Held back	-0.208	0.008	-25.121	0.000	-0.224	-0.192
Lower-middle SES	0.061	0.014	4.428	0.000	0.034	0.088
Middle SES	0.230	0.016	14.621	0.000	0.199	0.261
Upper-middle SES	0.398	0.019	20.867	0.000	0.361	0.436
Upper SES	0.579	0.038	15.270	0.000	0.504	0.653
Subsidised independent	0.051	0.011	4.521	0.000	0.029	0.073
Independent	0.023	0.035	0.657	0.511	-0.046	0.092
Interaction effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Prior attainment & Male	-0.060	0.003	-18.330	0.000	-0.067	-0.054
Male & Held back	0.021	0.010	2.022	0.043	0.001	0.041
Prior attainment & Held back	-0.069	0.006	-12.460	0.000	-0.080	-0.058
Prior attainment & Lower-middle SES	-0.020	0.006	-3.145	0.002	-0.032	-0.007
Prior attainment & Middle SES	-0.012	0.007	-1.602	0.109	-0.026	0.003
Prior attainment & Upper-middle SES	-0.010	0.009	-1.114	0.265	-0.028	0.008
Prior attainment & Upper SES	-0.069	0.021	-3.268	0.001	-0.110	-0.027
Prior attainment & Subs. independent	-0.001	0.005	-0.107	0.915	-0.010	0.009
Prior attainment & Independent	0.048	0.020	2.464	0.014	0.010	0.087

† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.

On another front, school SES has a distinct effect on progress in Spanish from that on Mathematics, when making this variable interact with prior attainment. Pupils in non-low SES schools are not expected to make more progress than pupils in low-SES schools; on the contrary, pupils in lower-middle SES schools and upper-SES schools are expected to make significantly less progress; while pupils in middle and upper-middle SES schools do not significantly differ from them. Moreover, when fitting an interaction between prior attainment and institutional type of the school on Spanish Language, it is appreciated that only pupils in independent schools are expected to make more progress than the rest of the pupils in State-funded and subsidised independent schools. This is clearly different to the patterns recorded in the CVA model for Mathematics (Table 4.12), where pupils in subsidised independent schools were expected to make more progress, i.e. when making prior attainment and school type interact.

This seems to indicate that the institutional type of the schools seems to matter more in the case of progress in Spanish. Put another way, attending an independent school only represents an advantage in terms of progress in Spanish Language, but not in Mathematics. These patterns seem to be yet another reflection of the effect of the schools' socio-economic sorting in Chile. In table 4.18, the random part of the CVA for Spanish is presented.

Table 4.18: Random part of the full CVA model for progress in Spanish Language

Parameter	Estimate	S.E.	Correlation	95% Coverage Int.	
Level 4: Local authority					
Variance (Intercept)	0.002	0.001	--	0.001	0.004
Covariance (Intercept, Male)	0.001	0.0003	0.705	0.0000	0.001
Variance (Male)	0.0002	0.0002	--	-0.0002	0.001
Covariance (Intercept, Prior attainment)	0.001	0.0003	0.471	0.0001	0.001
Variance (Prior attainment)	0.001	0.0002	--	0.0004	0.001
Covariance (Prior attainment, Male)	0.0001	0.0002	0.123	-0.0002	0.0003
Level 3: School					
Variance (Intercept)	0.025	0.001	--	0.022	0.028
Covariance (Intercept, Male)	0.003	0.001	0.471	0.001	0.004
Variance (Male)	0.001	0.001	--	-0.0001	0.003
Covariance (Intercept, Prior attainment)	0.001	0.0004	0.195	0.00004	0.002
Variance (Prior attainment)	0.001	0.0003	--	0.0003	0.001
Covariance (Prior attainment, Male)	-0.0003	0.0003	-0.313	-0.001	0.0003
Level 2: Classroom					
Variance (Intercept)	0.019	0.001	--	0.017	0.021
Covariance (Intercept, Male)	0.002	0.001	0.217	0.001	0.004
Variance (Male)	0.006	0.001	--	0.003	0.009
Covariance (Intercept, Prior attainment)	0.004	0.0005	0.446	0.003	0.004
Variance (Prior attainment)	0.003	0.0004	--	0.003	0.004
Covariance (Prior attainment, Male)	0.002	0.001	0.374	0.001	0.003
Level 1: Pupil					
Variance (Intercept)	0.368	0.001	--	0.365	0.370

Regarding the random part of the CVA model for progress in Spanish, displayed in Table 4.18, some considerable differences are found with respect to the CVA model for Mathematics. In this model, it is found that the variance between pupils is larger (0.368 against 0.278) and the variance between schools is smaller (0.025 against 0.053). It is worth noting that all higher-level intercept variance estimates are smaller in this model for Spanish when compared to their counterparts for Mathematics. This would be indicating that progress in Spanish is more associated to unobserved heterogeneous pupil characteristics than progress in Mathematics.

Overall, the two models seem to be reflecting on profound inequalities that are found in the Chilean education system. It is clear that progress in both subjects seems to depend more on socio-economic characteristics of the schools and the pupils (non-malleable factors), rather than on what schools can actually modify to contribute to their pupils' progress. Given that the same trends are broadly found in both models and that scores in both subjects are known to be significantly and substantively associated, a multilevel model including both outcomes is highly relevant. This model will be further developed and analysed in Chapter 5.

4.5. Conclusions

In order to analyse the academic progress made by Chilean pupils in secondary schools in 2006, a 4-level random coefficients model has been implemented. Since the implemented models for attainment in Mathematics and later on in Spanish Language are conditional upon prior attainment and control for socio-economic variables at the level of pupils and schools, they can be regarded as contextualised value-added models. The main focus of this chapter has been set on the construction of the model for progress in Mathematics; however, most of the conclusions are applicable for the case of progress in Spanish, given that both univariate CVA models were implemented following the same theoretical and methodological principles discussed in Chapters 2 and 3.

The main conclusions arising from this chapter are:

a) The level of schools is not the only relevant source of variation in Mathematics progress (nor in Spanish Language), since variation between classrooms and local authorities is not negligible. The level of primary schools has also been found to be relevant; however, this is explored in more depth in Chapter 5.

b) There are a number of non-malleable educational, socio-economic and demographic characteristics at the level of pupils and schools that significantly affect pupils' progress in Mathematics (and Spanish Language) that need to be controlled for in order to implement a fair school accountability system in Chile.

c) Pupils' progress in Mathematics significantly varies across classrooms, schools and local authorities.

d) Differences between school accountability measures derived from diverse statistical models are not negligible.

Additionally, the full 4-level CVA model for Mathematics differs from the traditional 2-level school value-added models in two key aspects:

a) It assesses the variation from 4 levels, which have been shown to be highly significant and crucial to estimate more reliable school effects.

b) It effectively decreases the bias that arises in traditional value-added models controlling for prior attainment only (raw value-added models as described in Chapter 2), where prior attainment scores are highly correlated with school raw averages in subsequent attainment, which makes schools serving the best performers in previous tests obtain the highest value-added estimates.

Results from the CVA model for Mathematics reveal the extent of the inequalities of the Chilean education system; they show that progress in Mathematics is positively associated with living in a non-low income household, being male, attending a subsidised independent, non-low SES school. On the other hand, progress in Mathematics is negatively associated with living in a low-income household, being held back at least once during primary school, being a female, attending a low-SES school and attending a state-funded school. On the other hand, the gender gap manifests itself differently in Language, where female pupils are advantaged with respect to male pupils.

It is also found in the CVA model for Mathematics that male pupils progress more than female pupils in general, a tendency that is only remedied when (average) prior attainment is high in a particular school. Furthermore, it is found counter-intuitively that, in general, pupils in independent schools do not progress more in Mathematics than those in state-funded schools. However, progress must not be confused with attainment. Pupils in independent schools have not attained more than expected, given their socio-economic background and their prior attainment; in other words, they have not progressed more than pupils in State-funded schools. That is, independent schools do not seem to add value to their pupils' educational trajectories in Mathematics. With regard to progress in Language, results show a different pattern, where pupils in independent schools outperform all other pupils.

Nevertheless, a word of caution is needed on these results. Differences between schools' estimated value-added scores can partially be the result of a selection process in the school system. This selection process can be reflected upon a correlation between prior attainment and the CVA measures due to school choice based on school averages, which produces bias in the estimated random effects (Ebbes et al., 2004). This is an issue that is not directly addressed in this research, but it certainly requires more investigation. In this research, school selectivity is at least partially controlled for by specifying prior attainment and year repetition, as mentioned before.

In sum, the CVA model presented in this paper extends and improves the traditional 2-level school value-added models insofar as it explicitly assesses the variation between pupils, classrooms, secondary schools and local authorities. These levels have been shown to be highly significant and crucial to estimate more precise and informative school effects for accountability purposes mainly. More specifically, adjusting for the effect of local authorities allows contextualising the information about school performance within a given territory to which most often parents are limited to choose from and to which local government administrations are limited to intervene. On the other hand, adjusting for classroom effects allows unveiling differences within schools that traditional models obscure; such differences might be relevant for policy interventions and for parents who could make choices based on overall school effects rather than a mixture of effects proceeding from the schools and the classrooms. Parents cannot choose from the latter and authorities at the levels of schools and local government ought to identify within-school differences for intervention, and hence the relevance of distinguishing between the two. Although the focus has been set on external school accountability, these extensions to the traditional CVA models (mainly specifying further levels of variation) constitute non-trivial and non-negligible adjustments when feeding back to diverse stakeholders, because they take a more thorough account of the complexity of the school performance phenomenon.

In light of the apparent differences that are found in pupils' academic progress due to socio-economic inequalities, along with within and between-school differences and geographical effects, it is evident that a school accountability system that does not thoroughly consider them will undoubtedly be unfair. This is especially relevant when considering what the stakes are. As discussed in chapter 2, the new Chilean Education Quality Assurance System (Sistema Nacional de Aseguramiento de la Calidad de la Educación Escolar, SAC) foresees the utilisation of the school classifications for supporting schools with insufficient performance; however, persistent insufficient performance may result in closure (Agencia de Calidad de la Educación, 2014; San Martín & Carrasco, 2013). The choice of methods is, therefore, a key player in this high stakes game as seen in this chapter and should not be taken as a trivial matter.

Other theoretically possible extensions of this CVA model include assessing additional contextual random effects, such as neighbourhood and family, as well as the carry-over effects from primary schools. In this chapter, the effects of primary schools have indeed been found to be significant in a 5-level empty cross-classified model for attainment in Mathematics. However, further specifications go beyond the scope of this chapter, because they are not straightforward extensions to the standard multilevel model. In the case of the effects of neighbourhood, data are not available to analyse them.

Given the similarities between the CVA models for Mathematics and Spanish, as well as the relevant association between both subjects, and considering that primary schools have been shown to be a significant contribution to the model, a bivariate 5-level cross-classified CVA model is implemented next in Chapter 5 in order to disentangle the complexity of school-value added further.

Chapter 5: Analysing value-added for external school accountability with a bivariate multilevel model

5.1. Introduction

The main purpose of this chapter is to develop a more comprehensive, more reliable and ultimately fairer school value-added model for accountability, embodying a broader concept of school effectiveness in which academic progress is viewed as a multifaceted phenomenon. This chapter builds up from the conclusions of chapter 4, to integrate the following elements in a single CVA model: a) assessing the effects of set of widely known influential socio-economic and demographic characteristics of the pupils and the schools; b) specifying the additional levels of classrooms and local authorities as significant sources of variation, in addition to the traditional 2-level models of pupils nested within schools; c) extending the CVA model further to assess the variation between primary schools; and d) analysing the correlation between different subjects, i.e. Mathematics and Language.

The bivariate CVA model is built up sequentially, as described in the analytical strategy in Chapter 3. Relevant results arising from the implementation of each step are presented and discussed in detail throughout the chapter. Similar to how results are presented in Chapter 4, the construction of the bivariate CVA model of this chapter is described starting off with the variance components models (section 5.4); then, moving on to the raw CVA model controlling for prior attainment only (section 5.5) as a baseline, from which the influence of the characteristics of the schools and the pupils on progress in both subjects will be assessed (sections 5.6 and 5.7). Then, in sections 5.8 and 5.9, further insight is given into differential school effects that can only be estimated through this specification.

By the end of the chapter, in section 5.10, the practical implications of the full model are analysed by comparing school accountability measures derived from a traditional approach and from this novel specification. Finally, the conclusions arising from these analyses are discussed.

5.2. Exploring the relationship between Mathematics and Language test scores

In this section, the purpose is to explore the relationship between the scores in the standardised SIMCE tests of Mathematics and Spanish. As hinted in the previous chapters, achievement in the subjects of Mathematics and Spanish is reasonably believed to be associated.

Educational researchers have found that language attainment is positively associated with Mathematics achievement (see for example: Tate, 1997; Wang & Goldschmidt, 1999) along with other socio-economic, educational and demographic characteristics at the level of schools and pupils in models similar to those presented in the previous chapter. Nevertheless, a normal response univariate multilevel model cannot reliably account for this presumable correlation. Hypothetically speaking, if a model for performance in Mathematics is specified with a set of explanatory variables that include, for instance, socio-economic characteristics and language-related indicators, issues of multicollinearity and possibly endogeneity may arise. This is because such language proficiency indicators are also very likely to be associated with (or affected by) the socio-economic variables in the model.

As seen in the models of Chapter 4, progress in Mathematics and Spanish are both significantly associated with income, gender, year repetition, school institutional type and school SES, and hence in the event of specifying Spanish attainment as an additional predictor of Mathematics, spurious coefficients would be estimated, since attainment in the Spanish Language test could also be specified as a function of the other specified variables. In Figure 5.1 below, the relationship between the standardised scores in Mathematics and Language obtained by pupils in Year 10 in 2006 is explored.

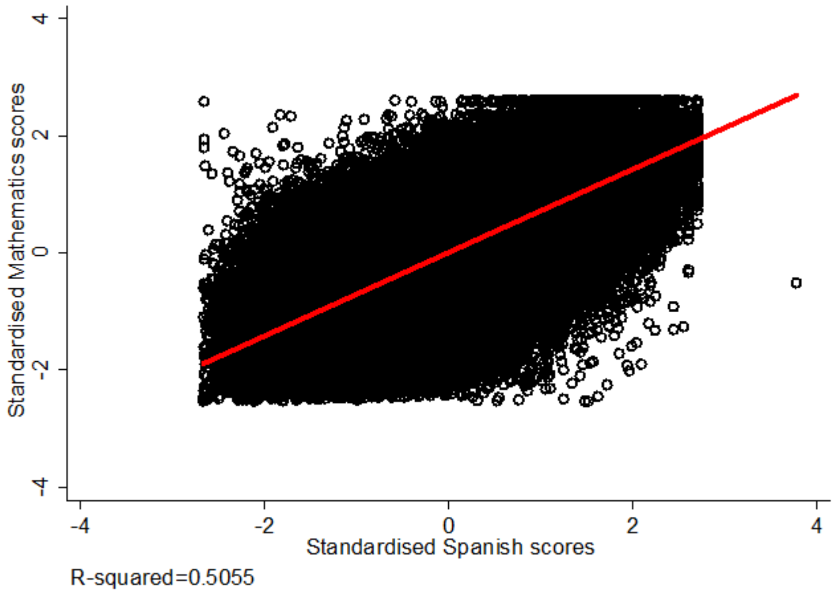


Figure 5.1: Relationship between Mathematics and Spanish Language standardised scores at the pupil level

In Figure 5.1, it can be appreciated that there is great variability at the pupil level, where many pupils with high scores in Mathematics recorded low scores in Spanish and vice versa. However, the overall picture, although not clear, is that the higher the score obtained in Mathematics, the greater the chance of obtaining high scores in Spanish. This is also demonstrated by a fairly high value for the R-square of 0.506 (correlation of 0.711). Not surprisingly, when averaging Mathematics and Spanish attainment at the level of schools, the pattern in the data becomes clearer as observed in the next figure.

In Figure 5.2, it is observed that there is a quite high correlation between the school average attainment in Mathematics and the school average attainment in Spanish. Thus, schools with a high average attainment in Mathematics are most likely to record a high average attainment in the Spanish Language test as well. The R-square coefficient in this case is 0.8903 (correlation of 0.944), which indicates that the school average of any of the tests is quite a good predictor of the school average in the other test.

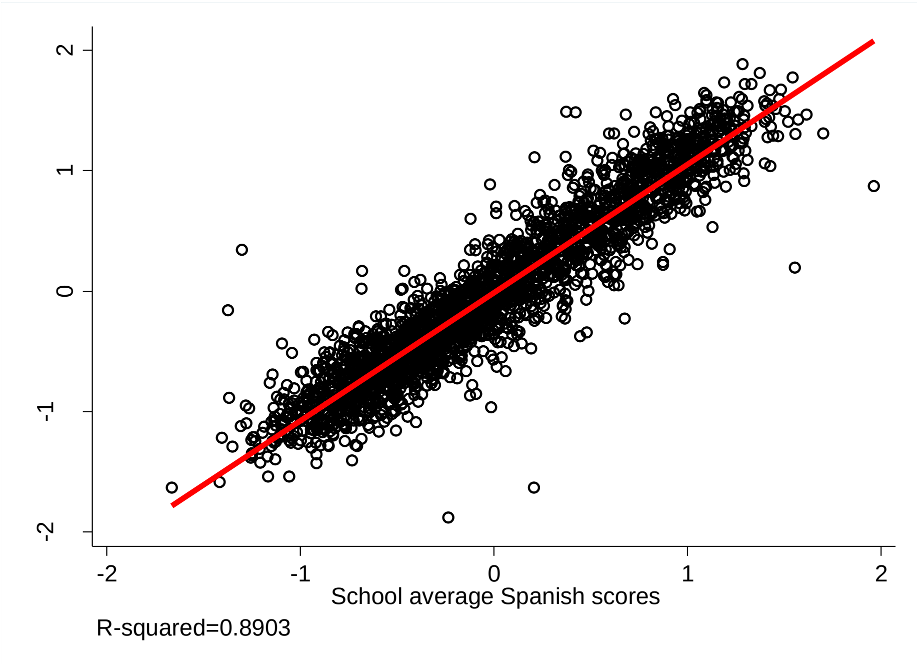


Figure 5.2: Relationship between Mathematics and Spanish Language standardised scores aggregated at the school level

As mentioned previously, a bivariate multilevel model is the best approach for taking into account this correlation without producing a model that suffers from unacceptable levels of multicollinearity or endogeneity. In the next section, this model is implemented as outlined in equations 12 and 13 of Chapter 3 and following the same 'bottom-up' analytical strategy outlined also in Chapter 3 and implemented for the univariate models presented in chapter 4.

5.3. Is there a relationship between progress and school value-added in Mathematics and Language?

In this section, the main aim is to ascertain whether there is an underlying relationship between Mathematics and Language. If proved significant, this relationship could be considered evidence in favour of the multidimensionality of school value-added and pupils' progress, manifested through attainment in various subjects, in this case Mathematics and

Language. To test the significance of the correlation between the two subjects, two bivariate multilevel models are fitted and compared: a) a baseline 2-level model with two outcomes whose correlations at both levels were constrained to be zero, and b) an unconstrained bivariate 2-level model with a full covariance matrix.

From a statistical point of view, another important advantage of using the multivariate specification is that the estimation of a joint model uses the information more efficiently (Goldstein, 2011). For instance, should there be pupils missing one test score but not the other, the covariance matrix could still be estimated efficiently using the available information on the other test score. In contrast, when specifying two separate models, missing information on one test score is simply disregarded. Furthermore, the multivariate specification allows testing hypotheses across outcomes; for instance, it is possible to test the significance of any explanatory variable on both outcomes jointly or separately, which can be of substantive interest for unveiling differential effects.

As observed in Table 5.1, the specification of an unconstrained bivariate multilevel model represents a significant improvement to the overall fit of the model (p<0.001), which implies that the Mathematics and Language test scores are significantly correlated. It is also appreciated that both sets of effects -fixed and random- remain mostly unchanged with this specification.

Table 5.1: Model comparison between constrained multilevel model with two uncorrelated outcomes and unconstrained bivariate multilevel model

	Parameters†	**Constrained**	**Unconstrained**
	Fixed Part	**Coef. (s.e.)**	**Coef. (s.e.)**
	Intercept Maths	0.018 (0.015)	0.018 (0.015)
	Intercept Language	0.027 (0.013)	0.024 (0.013)
Level	**Random Part**	**Coef.**	**Coef.**
Secondary Schools	Variance (Intercept Maths)	0.51	0.509
	Covariance (Language, Maths)	--	0.433
	Variance (Intercept Language)	0.392	0.392
Pupils	Variance (Intercept Maths)	0.569	0.569
	Covariance (Language, Maths)	--	0.347
	Variance (Intercept Language)	0.67	0.67
Model fit information	**-2*loglikelihood**	970233.877	888795.359
	AIC	970245.877	888811.359
	Number of parameters (k)	6	8
	Chi-square (X^2)	--	81438.518
	p(X^2)	--	<0.001
	N	202,605	202,605

† Parameters were obtained via IGLS estimation.

The comparison of these two models not only allows unveiling the statistical relevance of estimating a bivariate model, but also reveals that academic performance is not a unidimensional phenomenon, which is more important from a substantive point of view. This implies that underlying the observed test scores in Mathematics and Language; there exist other mechanisms and relationships that influence academic performance. In other words, subjects are not learned by pupils in an isolated way and subjects are not taught (intentionally or unintentionally) in an isolated way either. Furthermore, had the data been available, other relationships between different subjects could have been found.,

Using the covariance and variance terms of the unconstrained model, it is possible to estimate the correlation of the Mathematics and Language scores at the level of secondary schools and pupils. At the secondary school level, it is found that this correlation is 0.969, which means for instance that a secondary school achieving high scores in Mathematics on average would also be expected to score high on average in the Language test. In contrast, at the pupil level this relationship is not so clear, but also fairly high (0.562). Having demonstrated that the specification of a bivariate multilevel model is indeed fruitful, this empty model is extended further with additional levels of variation to test whether the bivariate multilevel model benefits from increasing its complexity.

5.4. Decomposing the variation in Mathematics and Language test scores

The first step in analysing progress in Mathematics and Spanish Language jointly, is to implement the empty models to assess the main significant sources of variation in the data in the same way as seen for the univariate CVA models in Chapter 4. Following the analytical strategy as outlined in Chapter 3, the significance of the specification of additional levels to the basic 2-level (pupils nested within schools) bivariate model was assessed by performing the likelihood ratio test and monitoring the reduction of the value of the Akaike's Information Criterion (AIC).

Furthermore, the addition of the level of primary schools requires using MCMC estimation (as explained in Chapter 3, section 3.7.2), which makes it necessary to use alternative goodness of fit measures. As mentioned in Chapter 3, the Deviance Information Criterion (DIC) is a generalisation of the AIC used in multilevel models fitted via MCMC. In this case, the values of the DIC of the models without cross-classification (2, 3 and 4 levels) were compared to the value of the DIC of the model with cross-classification (5 levels). In such comparison, a lower value of the DIC diagnostic for the more complex models in comparison to the less complex models, indicates a better fit (Browne, 2012). In Table 5.2, the results for the 2, 3, 4 and 5-level empty bivariate models are presented.

Results from the empty bivariate models show that the addition of every level of variation is a significant contribution to the overall fit of the model. The DIC diagnostic measure is drastically reduced with the specification of the class level (difference between DIC's is 24,870.16), while this is not as pronounced for the case of the addition of the local authority level (difference is 4.062), but still a relevant improvement given that the effective number of parameters (pD) is also reduced because of this specification. These results are also consistent with what is found for the 2, 3 and 4-level models (which can be fitted via IGLS estimation), where values of the AIC drop with every additional level specified. Finally, the addition of the level of primary schools, given its cross-classified structure, increases the overall complexity of the model with a larger number of effective parameters; however, the DIC indicates that this is a significant trade-off, with a large reduction of the deviance (difference with respect to 4-level model is 2,397.701) and hence a better fit.

Table 5.2: Summary of the empty bivariate models for attainment in Mathematics and Language

	Parameter†	**2 levels‡**	**3 levels**	**4 levels**	**5 levels**
	Fixed Part	**Coef. (s.e.)**	**Coef. (s.e.)**	**Coef. (s.e.)**	**Coef. (s.e.)**
	Intercept Maths	0.012 (0.015)	0.002 (0.015)	-0.137 (0.023)	-0.189 (0.029)
	Intercept Language	0.018 (0.013)	0.007 (0.013)	-0.122 (0.021)	-0.172 (0.027)
Level	**Random Part**	**Coef.**	**Coef.**	**Coef.**	**Coef.**
Secondary Schools	Var. (Intercept Maths)	0.51	0.47	0.396	0.366
	Cov. (Language, Maths)	0.433	0.402	0.334	0.307
	Var (Intercept Language)	0.393	0.362	0.3	0.275
Pupils	Var. (Intercept Maths)	0.569	0.502	0.502	0.492
	Cov. (Language, Maths)	0.347	0.292	0.292	0.284
	Var. (Intercept Language)	0.67	0.615	0.616	0.606
Classes	Var. (Intercept Maths)	--	0.095	0.095	0.093
	Cov. (Language, Maths)	--	0.077	0.077	0.076
	Var. (Intercept Language)	--	0.076	0.077	0.075
Local Authorities	Var. (Intercept Maths)	--	--	0.069	0.073
	Cov. (Language, Maths)	--	--	0.063	0.067
	Var. (Intercept Language)	--	--	0.058	0.062
Primary Schools	Var. (Intercept Maths)	--	--	--	0.015
	Cov. (Language, Maths)	--	--	--	0.013
	Var. (Intercept Language)	--	--	--	0.015
Model fit	DIC	882,639.635	857,769.5	857,765.4	855,367.7
	pD	5,048.381	11,476.94	11,470.71	14,617.81
	N	202,605	202,605	202,605	202,605

† Obtained via MCMC, using IGLS estimates as starting values. Chain length: 5,000; burn-in: 500, storing all iterations.
‡ Estimates from Table 5.1 differ from Table 5.2 because of the different estimation procedures utilised.

The variance estimates at the school and pupil level differ quite considerably across the models, where the between-school variance in Spanish is much lower than for the case of Mathematics, which would be indicating that performance in Spanish is more associated with pupil characteristics and abilities than performance in Mathematics.

On another front, as seen in the previous chapter, it can be easily noted that the school effects (school-level variance) are clearly overestimated in a basic 2-level model (pupils nested within secondary schools). In both subjects, the overestimation of the school effects ranges between 28% and 30%, when comparing the variance estimates of the 2-level model with the 5-level cross-classified model. In Table 5.3, the variance partitioning is presented in detail.

According to the variance partition coefficients, the variance at the higher levels is consistently larger across all the empty models for the case of the Mathematics test in comparison to the results in Spanish. Conversely, the variance due to the pupil level is higher in the Spanish test, which may be indicating that attainment in Language is more associated with pupil characteristics and abilities than attainment in Mathematics. It can also be observed that the carry-over effects from primary schools are comparatively small (although significant) with respect to the effects from secondary schools. However, amongst the four empty models, the most important (largest) source of variation is always the pupil level. These results are consistent with what has been found in the previous chapter: that is, the variance components are broadly similar to those of the univariate models in Chapter 4.

Table 5.3: Variance partition coefficients of the empty bivariate models for attainment in Mathematics and Language

Level	Test	2 levels	3 levels	4 levels	5 levels
Secondary Schools	**Mathematics**	47.27%	44.05%	37.29%	35.23%
	Language	36.97%	34.38%	28.54%	26.62%
Pupils	**Mathematics**	52.73%	47.05%	47.27%	47.35%
	Language	63.03%	58.40%	58.61%	58.66%
Classes	**Mathematics**	--	8.90%	8.95%	8.95%
	Language	--	7.22%	7.33%	7.26%
Local Authorities	**Mathematics**	--	--	6.50%	7.03%
	Language	--	--	5.52%	6.00%
Primary Schools	**Mathematics**	--	--	--	1.44%
	Language	--	--	--	1.45%
Total	**Mathematics**	**100%**	**100%**	**100%**	**100%**
	Language	**100%**	**100%**	**100%**	**100%**

Since the specification of additional levels to the basic model structure of pupils nested within schools has proved to be fruitful, the upcoming bivariate analyses are fitted accordingly. On another front, these empty bivariate models cannot be considered as value-added models in the rightful sense, since they do not control for any measure of prior attainment. The next model to be fitted is the raw value-added bivariate model, where prior attainment in both subjects is specified as the only explanatory variable.

5.5. Controlling for prior attainment in Mathematics and Language

The inclusion of prior attainment, as expected, produces drastic reductions of the random effects at all levels, when compared to the empty models. It can be appreciated that both measures of prior attainment have very similar effect on their corresponding measures of subsequent attainment and these are consistent with the univariate models described in Chapter 4. This raw value-added bivariate model can be understood as an extension to the multivariate case of the concept of "type AA value added model" developed by Timmermans et al. (2011) for the univariate case.

It is found here that a one standard deviation increase of prior attainment in Mathematics produces a 0.564 standard deviations increase in subsequent Mathematics tests; whereas a one standard deviation increase of the score in the first occasion of the Spanish test produces a 0.575 standard deviations increase in the second occasion of the Spanish Language test.

Table 5.4: Raw value-added bivariate model for progress in Mathematics and Language

	Parameter†	**Coef.**	**S.E.**	**95% C. I.**	
	Fixed Part				
	Intercept Maths	-0.057	0.012	-0.079	-0.034
	Intercept Spanish	0.564	0.002	0.561	0.567
	Maths prior attainment	-0.050	0.010	-0.070	-0.030
	Language prior attainment	0.575	0.002	0.572	0.578
Level	**Random Part**	**Coef.**	**S.E.**	**95% C. I.**	
Secondary Schools	Variance (Intercept Maths)	0.120	0.004	0.111	0.129
	Covariance (Language, Maths)	0.089	0.003	0.083	0.096
	Variance (Intercept Language)	0.077	0.003	0.071	0.083
Pupils	Variance (Intercept Maths)	0.285	0.001	0.283	0.287
	Covariance (Language, Maths)	0.082	0.001	0.080	0.083
	Variance (Intercept Language)	0.374	0.001	0.372	0.377
Classes	Variance (Intercept Maths)	0.040	0.001	0.038	0.042
	Covariance (Language, Maths)	0.026	0.001	0.024	0.027
	Variance (Intercept Language)	0.026	0.001	0.025	0.028
Local Authorities	Variance (Intercept Maths)	0.016	0.003	0.011	0.021
	Covariance (Language, Maths)	0.013	0.002	0.009	0.018
	Variance (Intercept Language)	0.013	0.002	0.009	0.017
Primary Schools	Variance (Intercept Maths)	0.009	0.0004	0.008	0.009
	Covariance (Language, Maths)	0.005	0.0003	0.004	0.006
	Variance (Intercept Language)	0.005	0.0003	0.004	0.005

† Obtained via MCMC estimation. Chain length: 5,000; burn-in: 500; storing all iterations.

From this model a pseudo R-squared estimate can be computed from the residual variances at each level as seen in Hox (2010). Thus, prior attainment is found to have explained or accounted for 67.92% of the variance at the school level in Mathematics and 72.89% of the variance at the school level in Language; whereas at the pupil level, prior attainment explains 42.42% and 38.59% of the variance in Mathematics and Language, respectively.

Unsurprisingly, this bivariate raw value-added model represents a highly significant improvement of the overall fit of the model with respect to the variance components model, as seen in Table 5.5, where it is observed that a very large reduction of the DIC indicator occurs after controlling for prior attainment. The difference between the DIC value of the variance components model and DIC of the model controlling for prior attainment only is 142,471.449.

Table 5.5: Model fit comparison between bivariate variance components model and raw value-added bivariate model

Parameter†	**Variance components model (m0)**	**Raw value-added bivariate model (m1)**
DIC	761,098.419	618,626.970
pD	12,943.107	12,218.847
DIC (m0) - DIC (m1)	--	142,471.449
Units: Pupils	180,575	180,575
Units: Classrooms	7,459	7,459
Units: Primary schools	5,537	5,537
Units: Secondary schools	2,438	2,438
Units: Local Authorities	320	320

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iteration.

Notwithstanding, the 'explanatory' power of prior attainment is not surprising and it should be interpreted with caution, because this measure is considered to be only a control variable to measure progress in both subjects. Additionally, as it occurred with the raw value-added model for progress in Mathematics (Chapter 4), this raw value-added bivariate model suffers from being severely misspecified and biased towards overestimating the effects of those schools serving the highest achievers in previous tests. The next models control for a series of socio-economic, demographic and contextual variables at the school and pupil level in order to estimate more reliable, fair and accurate school effects.

5.6. How much of the progress made in Mathematics and Language is explained by the characteristics of the pupils?

The next stage of the modelling process, as outlined in the analytical strategy presented in Chapter 3, involves adding the explanatory variables at the lowest level of the data structure, which in this case is the pupil level. The socio-economic, demographic and educational characteristics of the pupils have been incorporated to the bivariate model in addition to the measures of prior attainment for both subjects. All parameters have been freely estimated without constraints, and hence all variables are assumed to have different effects on both outcomes.

As this model incorporates socio-economic and demographic characteristics of the pupils, this model can be defined as a true contextualised value-added model. Additionally, given that the only controls are made at the level of the pupils, this model can also be understood as the extension to the bivariate case of the "type A value added model", as defined by Timmermans et al. (2011) and Raudenbush (2004).

As discussed in Chapter 4, given that significant interaction effects are found, the main effects of the fixed part of the bivariate model are not to be interpreted in their own right, but in combination with the variables with which they are interacted. Some noticeable differences between progress in Mathematics and Language are evident from the fixed effects of the bivariate CVA model controlling only for pupil characteristics. Full details are given in Table 5.6.

Table 5.6: Fixed effects of the contextualised value-added bivariate model of progress in Mathematics and Language controlling for pupil-level explanatory variables only

Mathematics					**Language**				
Main effects†‡	**Coef.**	**S.E.**	**95% C.I.**		**Main effects**	**Coef.**	**S.E.**	**95% C.I.**	
Intercept	-0.080	0.011	-0.102	-0.057	Intercept	-0.036	0.009	-0.055	-0.018
Prior attainment	0.557	0.003	0.552	0.562	Prior attainment	0.602	0.003	0.597	0.608
Male	0.081	0.004	0.073	0.090	Male	-0.029	0.005	-0.038	-0.020
Lower-middle income	0.022	0.005	0.013	0.031	Lower-middle income	0.041	0.005	0.031	0.051
Upper-middle income	0.044	0.007	0.030	0.057	Upper-middle income	0.085	0.008	0.068	0.101
High income	0.073	0.010	0.053	0.093	High income	0.111	0.011	0.090	0.134
Held back	-0.275	0.008	-0.291	-0.258	Held back	-0.221	0.009	-0.238	-0.202
Interaction effects	**Coef.**	**S.E.**	**95% C.I.**		**Interaction effects**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Male	-0.009	0.003	-0.014	-0.003	Prior att. & Male	-0.050	0.003	-0.057	-0.044
Prior att. & Low-mid inc.	0.012	0.003	0.005	0.018	Prior att. & Low-mid inc.	-0.007	0.004	-0.014	-0.0002
Prior att. & Up-mid inc.	0.018	0.005	0.009	0.028	Prior att. & Up-mid inc.	0.006	0.006	-0.005	0.017
Prior att. & High income	0.013	0.006	0.002	0.024	Prior att. & High income	0.002	0.007	-0.011	0.015
Prior att. & Held back	-0.075	0.005	-0.085	-0.065	Prior att. & Held back	-0.072	0.005	-0.082	-0.061
Male & Low-mid income	-0.014	0.006	-0.026	-0.002	Male & Low-mid income	-0.004	0.007	-0.018	0.009
Male & Up-mid income	-0.020	0.009	-0.039	-0.002	Male & Up-mid income	-0.014	0.011	-0.034	0.007
Male & High income	-0.001	0.011	-0.023	0.021	Male & High income	0.022	0.012	-0.003	0.046
Male & Held back	0.082	0.009	0.065	0.100	Male & Held back	0.025	0.010	0.005	0.045
Low-mid inc. & Held back	0.006	0.011	-0.015	0.028	Low-mid inc. & Held back	0.007	0.012	-0.017	0.032
Up-mid inc. & Held back	0.007	0.019	-0.031	0.045	Up-mid inc. & Held back	-0.005	0.022	-0.048	0.038
High income & Held back	0.007	0.024	-0.041	0.055	High income & Held back	-0.029	0.027	-0.080	0.025

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.
‡ Reference categories: Female; Low income and Not held back.

First of all, the gender gap manifests itself differently in both outcomes as also noted in Chapter 4. Overall, female pupils are expected to make more progress than males in Language;

however, this relationship is moderated by prior attainment and year repetition. With regard to year repetition, boys who have been made to repeat are better off than female pupils, but worse off when considering prior attainment in Language. In Mathematics, on the other hand, boys are expected to score higher overall, but this effect is moderated by income, prior attainment and year repetition; boys are expected to make slightly less progress than girls when considering their prior attainment and also when they come from middle-income households. Nevertheless, male pupils who have been made to repeat at least one year in primary school are much better off than females, in terms of their progress in Mathematics. The harmful effect of year repetition is found in both subjects and is found to be larger for girls than boys; however, in Language progress it is less pronounced.

Another noticeable difference between both subjects in the fixed part of the model is that the effect of prior attainment on Language does not vary across all household income groups. It is only pupils in lower-middle income household who are expected to have slightly less progress than pupils in low-income households, whereas all other pupils follow the overall pattern, i.e. they progress as expected by their income group. In the case of gender, its effect on Language does not vary across income. This is clearly distinct from what is recorded in the case of Mathematics, where the effects of prior attainment and gender both vary significantly across income groups. Not only do pupils in non-low income households start off better in terms of attainment, but they also make more progress in Mathematics than pupils in low-income households. With regard to the effect of gender on Mathematics, it is found that male pupils in middle-income households (lower and upper) make significantly less progress than their female counterparts.

Even though the interaction between gender and income is only significant in one of the outcomes of the model, this effect will be fully kept in subsequent models to make the results in both outcomes more easily comparable. On the other hand, the interaction between pupils' retention and family income in both outcomes did not turn out to be significant and hence it will be suppressed from subsequent models. As for the random effects of this model, Table 5.7 shows the full details.

Unsurprisingly, the variance estimates of the intercepts of both outcomes have considerably dropped from what was estimated in the unconditional model (Table 5.2). This is very much expected, since the explanatory power of the socio-economic background on academic achievement has been thoroughly analysed and demonstrated in previous educational research. The drop in the variance estimates implies that once the pupils' socio-economic, demographic and educational characteristics have been taken into account, little variation between units can be found at all levels. This is especially true for the case of the levels of secondary schools and local authorities, where the variance estimates of the intercepts of both outcomes have been reduced by more than 70%; moreover, the variance estimate of the intercept of Language at the level of primary schools has also been reduced by more than 70%.

Table 5.7: Random effects of the contextualised value-added bivariate model of progress in Mathematics and Language controlling for pupil-level explanatory variables only

Level	Parameters†	Coef.	S.E.	Correlation	95% C.I.	
Secondary schools	Var. (Intercept Maths)	0.109	0.004	--	0.101	0.117
	Cov. (Int. Maths, Int. Language)	0.076	0.003	0.924	0.070	0.082
	Var. (Intercept Language)	0.062	0.003	--	0.057	0.068
Classes	Var. (Intercept Maths)	0.039	0.001	--	0.037	0.041
	Cov. (Int. Maths, Int. Language)	0.024	0.001	0.783	0.023	0.026
	Var. (Intercept Language)	0.025	0.001	--	0.023	0.026
Pupils	Var. (Intercept Maths)	0.279	0.001	--	0.277	0.281
	Cov. (Int. Maths, Int. Language)	0.080	0.001	0.250	0.079	0.082
	Var. (Intercept Language)	0.370	0.001	--	0.367	0.372
Local Authorities	Var. (Intercept Maths)	0.011	0.002	--	0.008	0.016
	Cov. (Int. Maths, Int. Language)	0.009	0.002	0.966	0.006	0.012
	Var. (Intercept Language)	0.007	0.001	--	0.005	0.010
Primary schools	Var. (Intercept Maths)	0.008	0.0004	--	0.007	0.009
	Cov. (Int. Maths, Int. Language)	0.004	0.0003	0.766	0.004	0.005
	Var. (Intercept Language)	0.004	0.0004	--	0.004	0.005

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.

Regarding the relationship between the Mathematics and Language intercepts, it is appreciated that there is a noticeable high correlation at the levels of Secondary schools and Local Authorities. This is expected to some extent given that these are high levels of data aggregation. These high correlations imply that a secondary school that is highly effective for its pupils in Mathematics is also expected to be highly effective in Language and vice versa. With the necessary modifications, the same is true for Local Authorities. Although not as high as for the case of the aggregated levels of local authorities and secondary schools, the intercepts of both outcomes are also highly correlated at the levels of classes and primary schools, meaning that classes and primary schools with high or low levels of achievement in Mathematics are also similarly effective in Language attainment. Lastly, it is apparent from Table 5.6 that pupils are more heterogeneous because high (or low) achievement in Mathematics is not necessarily accompanied by high (or low) achievement in Language, which is implied by the weak correlation estimated between the intercept of Mathematics and Language at the pupil level.

This apparent inconsistency between the correlations between Mathematics and Language at different levels can be explained by the underlying traits that these intercepts represent. At the higher levels (i.e. classes, primary schools, secondary schools and local authorities), the intercepts of both outcomes represent a theoretical overall degree of comparative effectiveness of the particular higher level unit. For instance, a secondary school that puts in place a series of policies to boost its pupils' performance will most likely obtain good results in standardised tests. That is because such policies will most likely not be circumscribed to only one particular subject, but other subjects as well. This would point towards an overall effectiveness of that particular secondary school that would be allegedly higher than the overall effectiveness of another secondary school that does not put in place such policies. Given that most differences between secondary schools, after controlling for non-malleable pupils' characteristics, will be found amongst the particular policies that they do and/or do not display. The most expected result is that the most effective secondary schools implement more comprehensive policies and that the least effective schools either implement ineffectively atomised policies or do not implement any special policy at all.

On the other hand, at the level of pupils, the weak correlation between the Mathematics and Language intercepts is probably a reflection of individual processes that are more related to intellectual or emotional abilities or even family characteristics. This is in addition to socio-economic and demographic characteristics, which are already controlled for in the model,. Thus, for example, pupils may be more inclined to develop further their skills in one of the subjects and leave the other aside according to what their parents or carers have directly or indirectly encouraged in earlier life. Other plausible causes might be associated with specific cognitive abilities that allow certain pupils to develop several skills in different domains, whereas other pupils may not have these innate abilities.

Table 5.8: Model fit comparison between raw value-added bivariate model and model with pupil-level explanatory variables only

Parameter†	**Raw value-added bivariate model (m1)**	**Pupil explanatory variables only (m2)**
DIC	618,626.970	615,020.807
Eff. No of parameters (pD)	12,218.847	11,968.430
DIC (m1) - DIC (m2)	--	3,606.162
Units: Pupils	180,575	180,575
Units: Classes	7,459	7,459
Units: Primary schools	5,537	5,537
Units: Secondary schools	2,438	2,438
Units: Local Authorities	320	320

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.

As expected, in Table 5.8 above, it can be appreciated that the inclusion of pupil-level explanatory variables has produced a highly significant improvement to the overall fit of the model when comparing to the previous model that controlled for prior attainment only (Table 5.4). It is observed here that the difference between the DIC value of the raw value-added bivariate model (m1, controlling for prior attainment only) and the DIC of the CVA model controlling for pupil characteristics is 3,606.162, which can be considered as a highly significant difference.

Although this specification proved to be a significant improvement to the overall fit, there is a known issue where school characteristics may still affect pupils' performance and progress, as seen previously in Chapter 4. Since what it is pursued in this analysis is the least biased estimation of the malleable school effects, the school characteristics that the model needs to control for are those that cannot be modified by the school. This is in addition to those compositional effects (pupil-level characteristics) that have already been specified in this bivariate CVA model controlling only for pupil-level explanatory variables. A detailed description of this step is presented in the next section.

5.7. Controlling for school context and characteristics

Between-school variation can theoretically be accounted for by various sources, amongst which it is possible to find:

a) the compositional effects, which have been analysed in the previous section;

b) the malleable school factors, which include all resources and policies put in place by the school in particular that boost or moderate pupils' performance; and finally,

c) the non-malleable school factors, which refer to the school context and characteristics over which the school has no power to modify or are intrinsically part of how the school is defined as an institution.

In this part of the analysis, the non-malleable factors to be incorporated into the models are the institutional type and the socio-economic status of the school. As described in Chapter 3, institutional type is an administrative denomination defined by the Chilean Ministry of Education to classify primary and secondary schools. According to the Chilean legislation, State-funded schools are comprehensive institutions owned and managed by the State through the Local Education Authorities, more specifically the Municipalities; Subsidised Independent schools are comprehensive institutions owned and managed by private parties, but partially funded as well as supervised by the Ministry of Education; Independent schools are selective schools fully funded by private parties and partially supervised by the Ministry of Education. On the other hand, School SES is an index elaborated by the Chilean Ministry of Education that takes into account the parents' level of qualifications, family income and a "vulnerability" indicator aggregated at the school level. This model corresponds to a "type B" value-added model as defined by Raudenbush (2004) and Timmermans et al. (2011), but extended to the bivariate case.

Table 5.9: Fixed-effects parameters of the random-intercepts bivariate CVA model of progress in Mathematics and Language controlling for pupil-level and school-level explanatory variables.

Fixed effects Mathematics†					**Fixed effects Language†**				
Pupil-level main effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level main effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Intercept	-0.343	0.015	-0.373	-0.315	Intercept	-0.243	0.012	-0.267	-0.221
Prior attainment	0.556	0.003	0.550	0.561	Prior attainment	0.600	0.003	0.595	0.606
Male	0.082	0.004	0.074	0.090	Male	-0.029	0.005	-0.038	-0.020
Low-mid income	0.018	0.004	0.009	0.026	Low-mid income	0.033	0.005	0.023	0.043
Up-mid income	0.032	0.007	0.018	0.045	Up-mid income	0.064	0.008	0.048	0.080
High income	0.046	0.010	0.027	0.066	High income	0.063	0.012	0.040	0.086
Held back	-0.272	0.008	-0.287	-0.256	Held back	-0.220	0.008	-0.236	-0.204
School-level main effects§	**Coef.**	**S.E.**	**95% C.I.**		**School-level main effects§**	**Coef.**	**S.E.**	**95% C.I.**	
Subs. Indep. school	0.055	0.014	0.028	0.081	Subs. Indep. school	0.049	0.011	0.029	0.070
Independent school	0.028	0.044	-0.062	0.111	Independent school	0.051	0.034	-0.015	0.116
Low-mid school SES	0.119	0.020	0.079	0.157	Low-mid school SES	0.090	0.015	0.060	0.118
Middle school SES	0.335	0.021	0.294	0.375	Middle school SES	0.273	0.016	0.240	0.304
Up-mid school SES	0.573	0.027	0.519	0.625	Up-mid school SES	0.453	0.020	0.413	0.493
Upper school SES	0.764	0.048	0.674	0.862	Upper school SES	0.609	0.038	0.535	0.686
Pupil-level Interaction effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level Interaction effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Male	-0.008	0.003	-0.014	-0.003	Prior att. & Male	-0.050	0.003	-0.056	-0.044
Prior att. & Low-mid inc.	0.011	0.003	0.005	0.017	Prior att. & Low-mid inc.	-0.009	0.003	-0.015	-0.002
Prior att. & Up-mid inc.	0.017	0.005	0.007	0.026	Prior att. & Up-mid inc.	0.004	0.006	-0.006	0.015
Prior att. & High income	0.012	0.006	0.001	0.023	Prior att. & High income	0.001	0.007	-0.012	0.014
Prior att. & Held back	-0.074	0.005	-0.084	-0.064	Prior att. & Held back	-0.072	0.005	-0.082	-0.062
Male & Low-mid income	-0.013	0.006	-0.025	-0.001	Male & Low-mid income	-0.004	0.007	-0.017	0.009
Male & Up-mid income	-0.020	0.009	-0.038	-0.002	Male & Up-mid income	-0.014	0.011	-0.034	0.007
Male & High income	-0.001	0.011	-0.023	0.021	Male & High income	0.020	0.012	-0.004	0.044
Male & Held back	0.082	0.009	0.064	0.099	Male & Held back	0.024	0.010	0.005	0.044

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.
‡ Reference categories for pupil-level variables: Female; Low income and Not held back.
§ Reference categories for school-level variables: State-funded school and Low school SES.

Table 5.9 displays the estimated fixed-effects parameters of the bivariate CVA model of progress in Mathematics and Language controlling for pupil and school level explanatory variables. It is observed here that only minor changes have been recorded in the estimated main effects for the pupil-level variables with respect to the estimated parameters of the previous model, which only controlled for pupil-level variables. The most pronounced change in these parameters is recorded for upper-middle income children in Language, where the estimated fixed effect dropped from 0.111 to 0.063 (standard deviations). Another salient feature is that the intercepts for both outcomes are much lower, although this is expected given that this current model controls for additional variables.

Concerning the interactions at the pupil-level, it is noted that these parameters have only suffered very slight changes with respect to the previous model. Considering that this model also controls for the aggregated school-level variable socio-economic status, the small changes in the estimated fixed effects related to income, either main or interaction, imply that the parameters of this model are robust, and even though some overlapping is expected, issues around multicollinearity are not evident.

With regard to the estimated fixed-effects of the school-level variables, it is first observed that subsidised independent schools have a significantly higher effect on pupils' performance in Mathematics and Language than State-funded schools (the reference category); thus, on average pupils attending subsidised Independent schools are expected to score 0.055 and 0.049 standard deviations more in Mathematics and Language, respectively, than their peers in State-funded schools. Meanwhile, the effect of attending an independent school is not significantly different from zero, because the estimated 95% credible interval contains zero, which implies that pupils in independent schools do not progress more than pupils in State-funded schools. This can be considered as evidence that most of the gap between low-achieving and high-achieving pupils is accounted for by socio-economic disparities. In other words, although pupils in Independent schools are high achievers, their progress in Mathematics and Language from primary to secondary schools is only as expected given their socio-economic advantages in comparison to their socio-economically disadvantaged peers attending State-funded schools.

On the other hand, as expected school socio-economic status has a large effect on Mathematics and Language scores. All non-low SES schools record, on average, significantly higher scores in both outcomes than low-SES schools, judging by the 95% credible intervals for the estimated coefficients of the four dummy variables which do not contain zero. Although school SES is an ordinal variable, and hence, a linear effect cannot be estimated, an upwards trend is clearly appreciated. This trend shows that each higher school SES category has a larger difference with respect to the category of low-SES schools in both outcomes. Furthermore, none of the 95% confidence intervals for the estimated coefficients of each school SES category overlaps with any other, which implies that there is a clear distinction between schools according to their average socio-economic status. Thus, upper-SES schools are expected to increase their pupils' progress by 0.764 standard deviations more than the low-SES schools, whereas upper-middle SES schools are expected to have only 0.573 standard deviations more than low-SES schools in Mathematics. It is also noticed that differences between schools in Language according to SES are still remarkable but not as large as they are in Mathematics. For instance, upper-SES schools are expected to increase their pupils' performance in Language by 0.609 standard deviations more than low-SES schools, while they boost their pupils' performance in Mathematics even further by 0.764 standard deviations as seen previously.

Concerning the random effects of this model, changes in the estimated correlations between the intercepts at all levels have been recorded. The details are contained in Table 5.10.

In Table 5.10, it is appreciated that the correlation between the intercepts of Mathematics and Language at the level of secondary schools has decreased from the one recorded in the previous model (from 0.924 as recorded in Table 5.7 to 0.854). At the level of local authorities, a similarly moderate decrease is also appreciated (from 0.966 to 0.910), while at the rest of the levels, correlations remain mostly unchanged (differences are of 0.001 or less). Overall, this indicates that school characteristics contributed in explaining only a small proportion of the correlations recorded at each level.

Table 5.10: Random effects of the bivariate CVA model of progress in Mathematics and Language controlling for pupil and school level fixed effects only (m3)

Level	Parameters†	Coef.	S.E.	Correlation	95% C.I.	
Secondary schools	Var. (Intercept Maths)	0.058	0.003	--	0.053	0.063
	Cov. (Int. Maths, Int. Language)	0.036	0.002	0.854	0.033	0.039
	Var. (Intercept Language)	0.030	0.001	--	0.027	0.033
Classes	Var. (Intercept Maths)	0.039	0.001	--	0.037	0.041
	Cov. (Int. Maths, Int. Language)	0.024	0.001	0.782	0.023	0.026
	Var. (Intercept Language)	0.025	0.001	--	0.023	0.026
Pupils	Var. (Intercept Maths)	0.279	0.001	--	0.277	0.281
	Cov. (Int. Maths, Int. Language)	0.081	0.001	0.251	0.079	0.082
	Var. (Intercept Language)	0.370	0.001	--	0.367	0.372
Local Authorities	Var. (Intercept Maths)	0.005	0.001	--	0.003	0.008
	Cov. (Int. Maths, Int. Language)	0.003	0.001	0.910	0.002	0.005
	Var. (Intercept Language)	0.003	0.001	--	0.002	0.004
Primary schools	Var. (Intercept Maths)	0.008	0.000	--	0.007	0.009
	Cov. (Int. Maths, Int. Language)	0.004	0.000	0.758	0.004	0.005
	Var. (Intercept Language)	0.004	0.000	--	0.003	0.005

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.

Table 5.11, below, displays the model fit comparison between the previous model which controlled only for pupil-level explanatory variables (m2) and the current model which also included school characteristics (m3). Although this specification significantly improves the overall fit of the model, it is still unsatisfactory, since it makes the unrealistic assumption that pupil-level variables do not vary across schools. In this respect, it is particularly inaccurate to assume that there is no significant variation in prior attainment in both outcomes across schools. At the same time, gender has proved to have a rather distinct effect on each of the outcomes, which makes it plausible that differences between genders can be found across schools as well. The next model builds up on these ideas and parameterises prior attainment in Mathematics and Language, as well as gender, as random effects at the level of secondary schools.

Table 5.11: Model fit comparison between CVA bivariate model controlling for pupil characteristics only (m2) and CVA bivariate model controlling for pupil and school characteristics (m3)

Parameter†	Pupil characteristics only (m2)	Pupil and school characteristics (m3)
DIC	615,020.807	614,882.1
pD	11,968.430	11,832.79
DIC (m2) - DIC (m3)	--	135.64
Units: Pupils	180,575	180,575
Units: Classes	7,459	7,459
Units: Primary schools	5,537	5,537
Units: Secondary schools	2,438	2,438
Units: Local Authorities	320	320

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.

5.8. Does the relationship between prior and subsequent attainment in male and female pupils in Mathematics and Language vary across secondary schools?

Two sub-models were fitted in this stage of the analysis. The first of these sub-models specifies prior attainment in Mathematics and Language as random effects at the secondary school level. The second sub-model specifies being male as a random coefficient with a variance residual at the secondary school level as well as the previously fitted random coefficient for prior attainment in both subjects.

Theoretically, these random coefficients could have also been specified at the other higher levels (primary schools, classes and local authorities). However, the computational burden of such models is not compensated with the little additional knowledge that could have been gained from them. Since the focus of this modelling process is set at the secondary school level, the other higher levels will only contain random intercepts, but acknowledging that the existence of other underlying random processes is indeed plausible.

As observed in Table 5.12, the estimated fixed-effects parameters of this model closely match the estimated parameters from the previous model in Table 5.9.

There is no need to interpret the results from Table 5.12 any further than the interpretation provided for Table 5.9. The novelty of this extension to the model is the estimated random effects of pupil-level variables prior attainment in Mathematics and Language, as well as gender. This implies that the random part at the secondary school level contains not only the random intercepts of both outcomes and their corresponding covariance terms, but also the random coefficients for prior attainment in both subjects, as well as gender of pupils, alongside their corresponding covariance terms. The random part of the levels of pupils, classes, primary schools and local authorities remain unchanged from the previous model presented in Table 5.10, and hence these terms were suppressed from Table 5.13, for the sake of simplicity.

Table 5.12: Fixed-effects parameters of the random-coefficients bivariate CVA model of progress in Mathematics and Language controlling for pupil-level and school-level explanatory variables

Fixed effects Mathematics†					**Fixed-effects Language†**				
Pupil-level main effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level main effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Intercept	-0.337	0.015	-0.367	-0.309	Intercept	-0.240	0.012	-0.263	-0.218
Prior attainment	0.560	0.003	0.554	0.566	Prior attainment	0.603	0.003	0.596	0.609
Male	0.079	0.005	0.070	0.088	Male	-0.032	0.005	-0.041	-0.022
Low-mid income	0.017	0.004	0.008	0.026	Low-mid income	0.033	0.005	0.022	0.042
Up-mid income	0.031	0.007	0.018	0.045	Up-mid income	0.063	0.008	0.047	0.079
High income	0.044	0.010	0.026	0.065	High income	0.063	0.012	0.039	0.086
Held back	-0.270	0.008	-0.285	-0.255	Held back	-0.219	0.008	-0.235	-0.203
School-level main effects§	**Coef.**	**S.E.**	**95% C.I.**		**School-level main effects§**	**Coef.**	**S.E.**	**95% C.I.**	
Subs. Indep. school	0.049	0.015	0.022	0.080	Subs. Indep. school	0.045	0.011	0.023	0.067
Independent school	0.012	0.045	-0.077	0.096	Independent school	0.036	0.036	-0.036	0.106
Low-mid school SES	0.114	0.018	0.079	0.149	Low-mid school SES	0.086	0.014	0.059	0.116
Middle school SES	0.323	0.022	0.281	0.368	Middle school SES	0.263	0.017	0.231	0.296
Up-mid school SES	0.561	0.026	0.512	0.613	Up-mid school SES	0.444	0.020	0.404	0.485
Upper school SES	0.771	0.048	0.676	0.865	Upper school SES	0.616	0.039	0.540	0.694
Pupil-level interaction effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level interaction effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Male	-0.006	0.003	-0.012	0.000	Prior att. & Male	-0.049	0.003	-0.056	-0.042
Prior att. & Low-mid inc.	0.007	0.003	0.001	0.013	Prior att. & Low-mid inc.	-0.009	0.004	-0.016	-0.002
Prior att. & Up-mid inc.	0.011	0.005	0.001	0.020	Prior att. & Up-mid inc.	0.002	0.005	-0.009	0.013
Prior att. & High income	0.008	0.006	-0.004	0.020	Prior att. & High income	-0.001	0.007	-0.015	0.013
Prior att. & Held back	-0.070	0.005	-0.080	-0.060	Prior att. & Held back	-0.068	0.005	-0.078	-0.058
Male & Low-mid income	-0.013	0.006	-0.025	-0.001	Male & Low-mid income	-0.004	0.007	-0.018	0.010
Male & Up-mid income	-0.020	0.009	-0.038	-0.001	Male & Up-mid income	-0.014	0.011	-0.034	0.008
Male & High income	-0.003	0.012	-0.024	0.019	Male & High income	0.019	0.013	-0.007	0.044
Male & Held back	0.084	0.009	0.066	0.102	Male & Held back	0.028	0.010	0.008	0.048

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations. All fixed-effects parameters have an effective sample size (ESS) of at least 5,500. ESS values are larger than the chain length due to negative autocorrelation of some parameters. This unusual mixing of the random draws is likely to be associated with the estimation of effects across two different outcomes. This negative autocorrelation could eventually be remedied by increasing the thinning; however, this is not of primary interest and goes beyond the scope of this thesis.
‡ Reference categories for pupil-level variables: Female; Low income and Not held back.
§ Reference categories for school-level variables: State-funded school and Low school SES.

The first outstanding element of the random part in Table 5.13 is that the random intercepts and their corresponding covariance have not varied noticeably from what is recorded in the previous model. On another front, four residual variances have also been estimated, namely: the slope variances of prior attainment in Mathematics and Language, as well as the residual variances associated with being a male in both subjects. These additional residual variances are not interpreted in their own right, but with their associated covariance terms.

Table 5.13: Random effects at the secondary school level of the contextualised value-added bivariate model of progress in Mathematics and Language controlling for the random effects of pupil-level variables at the secondary school level (m4)

Parameters†	Coef.	S.E.	Correlation	95% C.I.	
Var. (Intercept Maths)	0.060	0.003	--	0.055	0.065
Var. (Intercept Language)	0.028	0.002	--	0.025	0.031
Cov. (Intercept Maths, Intercept Language)	0.036	0.002	0.873	0.032	0.039
Cov. (Prior att.Maths, Intercept Maths)	0.003	0.001	0.209	0.002	0.005
Cov. (Prior att. Maths, Intercept Language)	0.002	0.0005	0.210	0.001	0.003
Var. (Prior att. Maths)	0.004	0.0003	--	0.004	0.005
Cov. (Prior att. Language, Intercept Maths)	0.004	0.001	0.284	0.002	0.005
Cov. (Prior att. Language, Intercept Language)	0.003	0.0005	0.338	0.002	0.004
Cov. (Prior att. Language, Prior att. Maths)	0.003	0.0002	0.817	0.002	0.003
Var. (Prior att. Language)	0.003	0.0003	--	0.002	0.003
Cov. (Male Maths, Intercept Maths)	0.001	0.001	0.081	-0.001	0.003
Cov. (Male Maths, Intercept Language)	0.001	0.001	0.131	-0.00003	0.003
Cov. (Male Maths, Prior att. Maths)	-0.0002	0.0003	-0.035	-0.001	0.001
Cov. (Male Maths, Prior att. Language)	-0.0002	0.0003	-0.062	-0.001	0.0003
Var. (Male Maths)	0.005	0.001	--	0.003	0.006
Cov. (Male Language, Intercept Maths)	0.003	0.001	0.258	0.001	0.006
Cov. (Male Language, Intercept Language)	0.004	0.001	0.399	0.002	0.005
Cov. (Male Language, Prior att. Maths)	0.0001	0.0003	0.035	-0.001	0.001
Cov. (Male Language, Prior att. Language)	0.0004	0.0003	0.156	-0.0001	0.001
Cov. (Male Language, Male Maths)	0.002	0.0005	0.636	0.002	0.003
Var. (Male Language)	0.003	0.001	--	0.002	0.004

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations. Random effects at the levels of pupils, classes, primary schools and local authorities have been suppressed for clarity and because they remain mostly unchanged from the previous model (Table 5.10).

It can be appreciated that some of these covariance terms are not very informative in substantive terms and thus can be neglected as they are close to zero. Nevertheless, they are indeed statistically significant controls in the model. This is the case for some of the covariance estimates associated with being a male pupil in Mathematics, and more specifically the following: the covariance between the males' residual variance in Mathematics and the Mathematics overall intercept; the covariance between the males' residual variance in Mathematics and the Language overall Intercept; the covariance between the males' residual variance in Mathematics and the slope variance for prior attainment in Mathematics; and the covariance between the male pupils' residual variance in Mathematics and the slope variance for prior attainment in Language.

Other covariance estimates that are deemed negligible are associated with being a male pupil in Language, that is the covariance between the males' residual variance in Language and the slope variance for prior attainment in Mathematics. Finally, the covariance term between the male pupils' residual variance in Language and the slope variance residuals for prior attainment in Language is not too meaningful either.

All other covariance terms are deemed to have some potential to explore further the phenomenon of school performance, since they reveal how school policies can have diverse effects, depending on the subject and the pupils to whom they are applied. Overall, these covariance terms indicate that secondary schools that perform well in Mathematics also perform well in Language. A more thorough interpretation of these parameters is given after the full model is presented subsequently. This two-step stage in the modelling is statistically significant as Table 5.14 shows.

Table 5.14: Model fit comparison between CVA bivariate model controlling for pupil and school characteristics (m3) and the intermediate Random Coefficients CVA bivariate models (m4.1 and m4.2)

Parameter†	**Pupil and school characteristics (m3)**	**Random coefficients for prior attainment (m4.1)**	**Random coefficients for prior attainment and gender (m4.2)**
DIC	614,882.1	613,570.9	613,358.3
pD	11,832.79	12,878.65	13,339.34
DIC (A) - DIC (B)‡	--	1,311.17	212.63
Units: Pupils	180,575	180,575	180,575
Units: Classes	7,459	7,459	7,459
Units: Primary schools	5,537	5,537	5,537
Units: Secondary schools	2,438	2,438	2,438
Units: Local Authorities	320	320	320

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 5,000; burn-in: 500; storing all iterations.
‡ Where A is the less complex model and B is the model complex model.

Table 5.14 shows that in spite of the increase in the number of effective parameters (pD), both further extensions to the CVA bivariate model significantly improved the overall fit, regardless of the increased complexity. Given that the variance of prior attainment in both subjects and the variance associated with male pupils are not constant across secondary schools, a highly plausible hypothesis is that these effects also vary significantly according to secondary school characteristics. To explore this, it is necessary to specify cross-level interaction effects between prior attainment in both subjects and school type as well as school SES, and also between pupils' gender and these two school characteristics. The next section deals with this new extension to the model.

5.9. Does progress in Mathematics and Language vary according to secondary schools' characteristics?

As noted previously, this stage of the modelling process involves specifying cross-level interaction effects to investigate the following questions: a) Does progress in Mathematics vary according to school characteristics; b) Does progress in Language vary according to school characteristics?; c) Does male pupils' progress in Mathematics vary according to school characteristics; and d) Does male pupils' progress in Language vary according to school characteristics?

These school characteristics are denoted with the secondary school-level variables "school institutional type" and "school average SES", and hence including their interactions with the pupil-level variables entail the estimation of 24 additional parameters. Results obtained from an intermediate model showed that the interactions involving the gender of pupils and school characteristics did not turn out to be significant, and thus the 12 associated parameters were suppressed from the final model. Therefore, the full model includes all the estimated effects from the previous models as well as 12 additional parameters corresponding to the cross-level interactions between prior attainment in both subjects and the school characteristics. This is the final model in this process and so it was estimated using a longer MCMC chain (215,000 iterations) to ensure the robustness of the final estimates. Extended MCMC information from the full model can be found in Appendix 2.2.

There are a few outstanding features in Table 5.15; firstly, most of the pupil-level and school-level main effects have shown broadly similar values to those in the previous sections. However, the only pupil-level interaction effect that has remained significant is the interaction between pupils' gender and pupils' retention in both subjects, where it is observed that male pupils who have been made to repeat at least one year during primary school make more progress than their female peers in the same conditions (as noted also in Table 5.6). On the other hand, the interactions between prior attainment and self-reported average household income in both subjects as well as the interactions between pupils' gender and income have become non-significant. This is most likely due to the specification of the cross-level interaction between prior attainment and average school SES, which partially overlaps with the pupil-level variable income. Full details of this intermediate model can be found in Appendix 2.1.

Table 5.15: Fixed-effects parameters from the full CVA bivariate model for progress in Mathematics and Language, including cross-level interaction effects.

Fixed effects Mathematics†					**Fixed effects Language†**				
Pupil-level main effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level main effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Intercept	-0.342	0.016	-0.372	-0.312	Intercept	-0.234	0.012	-0.257	-0.210
Prior attainment	0.551	0.006	0.540	0.563	Prior attainment	0.615	0.006	0.604	0.626
Male	0.079	0.005	0.070	0.088	Male	-0.033	0.005	-0.042	-0.023
Low-mid income	0.016	0.004	0.008	0.025	Low-mid income	0.032	0.005	0.022	0.041
Up-mid income	0.032	0.007	0.018	0.046	Up-mid income	0.067	0.008	0.051	0.082
High income	0.044	0.010	0.025	0.062	High income	0.065	0.011	0.044	0.086
Held back	-0.269	0.008	-0.284	-0.254	Held back	-0.220	0.008	-0.236	-0.203
School-level main effects§	**Coef.**	**S.E.**	**95% C.I.**		**School-level main effects§**	**Coef.**	**S.E.**	**95% C.I.**	
Subs Indep. school	0.059	0.016	0.029	0.090	Subs Indep. school	0.049	0.012	0.026	0.072
Independent school	0.022	0.045	-0.067	0.110	Independent school	0.023	0.036	-0.048	0.094
Low-mid school SES	0.104	0.019	0.067	0.141	Low-mid school SES	0.071	0.015	0.043	0.100
Middle school SES	0.320	0.022	0.278	0.363	Middle school SES	0.254	0.017	0.222	0.287
Up-mid school SES	0.554	0.026	0.503	0.605	Up-mid school SES	0.436	0.020	0.396	0.475
Upper school SES	0.775	0.049	0.680	0.872	Upper school SES	0.622	0.039	0.544	0.700
Pupil-level interaction effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level interaction effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Male	-0.006	0.003	-0.012	0.0002	Prior att. & Male	-0.050	0.003	-0.056	-0.043
Male & Low-mid inc.	-0.013	0.006	-0.025	-0.001	Male & Low-mid inc.	-0.002	0.007	-0.016	0.012
Male & Up-mid inc.	-0.019	0.009	-0.038	-0.001	Male & Up-mid inc.	-0.014	0.011	-0.035	0.007
Male & High income	-0.002	0.012	-0.024	0.021	Male & High income	0.019	0.013	-0.006	0.044
Prior att. & Held back	-0.068	0.005	-0.078	-0.058	Prior att. & Held back	-0.069	0.005	-0.080	-0.058
Male & Held back	0.084	0.009	0.066	0.102	Male & Held back	0.029	0.010	0.009	0.049
Cross-level interaction effects‡§	**Coef.**	**S.E.**	**95% C.I.**		**Cross-level interaction effects‡§**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Subs. Indep.	0.012	0.005	0.002	0.022	Prior att. & Subs. Indep.	-0.001	0.005	-0.010	0.009
Prior att. & Indep. school	0.015	0.018	-0.020	0.050	Prior att. & Indep. school	0.033	0.019	-0.004	0.070
Prior att. & Low-mid SES	-0.010	0.006	-0.023	0.002	Prior att. & Low-mid SES	-0.023	0.006	-0.035	-0.011
Prior att. & Middle SES	0.022	0.007	0.008	0.036	Prior att. & Middle SES	-0.014	0.007	-0.028	-0.001
Prior att. & Up-Mid SES	0.023	0.009	0.005	0.040	Prior att. & Up-Mid SES	-0.010	0.009	-0.026	0.007
Prior att. & Upper SES	-0.016	0.019	-0.053	0.021	Prior att. & Upper SES	-0.052	0.020	-0.091	-0.013

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 215,000; burn-in: 500; storing all iterations. All fixed-effects parameters have an effective sample size of at least 3,800.
‡ Reference categories for pupil-level variables: Female; Low income and Not held back.
§ Reference categories for school-level variables: State-funded school and Low school SES.

From the cross-level interaction effects, some evidence is found that progress in Mathematics varies slightly according to schools' institutional type, where the interaction between prior attainment in Mathematics and subsidised independent school is significant. This implies that the slope of prior attainment in Mathematics is slightly steeper for pupils in subsidised independent schools in comparison to their peers in state-funded schools. On the other hand, there is no evidence that the slope of prior attainment is steeper for pupils in independent schools compared with pupils in State-funded schools (the reference category). In the case of Language, the interaction between prior attainment and school institutional type did not turn out to be significant, which means that the relationship between prior attainment (covariate) and subsequent attainment (outcome) in Language does not vary according to institutional type.

In the case of average school SES, evidence is found that the slope of prior attainment in Mathematics varies across SES levels, although this slope is not significant for all levels. Only pupils in middle SES and upper-middle SES schools make slightly more progress, i.e. the linear relationship between prior and subsequent attainment is steeper, than that of their peers in low-SES schools (the reference category), while pupils attending upper-SES schools do not seem to make more progress than pupils in low-SES schools. Meanwhile, no such evidence has been found for progress in Language, where none of these interactions are significant.

Regarding the random part of this full model, some variance and covariance estimates at the secondary school level have changed slightly from those presented in Table 5.13. Nonetheless this seeming instability is the result of a non-converged MCMC chain reported above, due to the difficulty of estimating such small random effects and the relatively limited length of the chain, which was not sufficient to produce an acceptable number of uncorrelated simulated random samples. Nevertheless, its length was indeed enough for model exploration and somewhat limited inference. These results are more reliable insofar as the MCMC chain has indeed converged after 215,000 iterations.

In Table 5.16, it can be observed that most of the random-effects parameters at the secondary school level are broadly similar to the ones reported in Table 5.13. Since the slope residual variances for the pupil-level variables male (gender) and prior attainment (in Mathematics and Language) cannot be interpreted in their own right, it is better to focus on the covariance terms, which are more easily interpreted using the estimated correlations.

Table 5.16: Random-effects parameters of the full random coefficients bivariate CVA model controlling for pupil and school level explanatory variables and cross-level interaction effects.

Levels	Parameters†	Coef.	S.E.	Correlation	95% C.I.	
Secondary Schools	Var. (Intercept Maths)	0.059	0.003	--	0.054	0.065
	Var. (Intercept Language)	0.028	0.002	--	0.025	0.031
	Cov. (Int. Maths, Int. Language)	0.035	0.002	0.873	0.032	0.039
	Cov. (Prior att. Maths, Int. Maths)	0.003	0.001	0.204	0.002	0.004
	Cov. (Prior att. Maths, Int. Language)	0.002	0.0005	0.214	0.001	0.003
	Var. (Prior att. Maths)	0.004	0.0003	--	0.003	0.005
	Cov. (Prior att. Language, Int. Maths)	0.003	0.001	0.288	0.002	0.005
	Cov. (Prior att. Language, Int. Language)	0.003	0.0005	0.344	0.002	0.004
	Cov. (Prior att. Language, Prior att. Maths)	0.003	0.0002	0.836	0.002	0.003
	Var. (Prior att. Language)	0.002	0.0003	--	0.002	0.003
	Cov. (Male Maths, Int. Maths)	0.002	0.001	0.111	-0.0004	0.004
	Cov. (Male Maths, Int. Language)	0.002	0.001	0.149	0.0001	0.003
	Cov. (Male Maths, Prior att. Maths)	0.00004	0.0003	0.010	-0.001	0.001
	Cov. (Male Maths, Prior att. Language)	-0.0001	0.0003	-0.044	-0.001	0.0005
	Var. (Male Maths)	0.004	0.001	--	0.003	0.006
	Cov. (Male Language, Int. Maths)	0.004	0.001	0.292	0.002	0.006
	Cov. (Male Language, Int. Language)	0.004	0.001	0.422	0.002	0.005
	Cov. (Male Language, Prior att. Maths)	0.0002	0.0003	0.054	-0.0005	0.001
	Cov. (Male Language, Prior att. Language)	0.0005	0.0003	0.182	-0.0001	0.001
	Cov. (Male Language, Male Maths)	0.002	0.001	0.647	0.001	0.003
	Var. (Male Language)	0.003	0.001	--	0.002	0.004
Classes	Var. (Intercept Maths)	0.039	0.001	--	0.037	0.041
	Cov. (Int. Maths, Int. Language)	0.025	0.001	0.784	0.023	0.026
	Var. (Intercept Language)	0.025	0.001	--	0.023	0.026
Pupils	Var. (Intercept Maths)	0.276	0.001	--	0.274	0.278
	Cov. (Int. Maths, Int. Language)	0.079	0.001	0.248	0.077	0.081
	Var. (Intercept Language)	0.368	0.001	--	0.365	0.370
Local Authorities	Var. (Intercept Maths)	0.004	0.001	--	0.002	0.006
	Cov. (Int. Maths, Int. Language)	0.003	0.001	0.858	0.001	0.004
	Var. (Intercept Language)	0.002	0.001	--	0.001	0.004
Primary Schools	Var. (Intercept Maths)	0.007	0.0004	--	0.007	0.008
	Cov. (Int. Maths, Int. Language)	0.004	0.0003	0.765	0.003	0.005
	Var. (Intercept Language)	0.004	0.0003	--	0.003	0.004

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 215,000; burn-in: 500; storing all iterations. All random-effects parameters have an effective sample size of at least 1,000.

Out of all these correlations, there are three very high correlations worth emphasising. Firstly, the correlation between the intercepts of Mathematics and Language (0.873) indicates that a secondary school performing well in Mathematics is highly likely to perform well in Language, and vice versa (see: Figure 5.3). Secondly, the slopes of prior attainment in Mathematics and Language are also unsurprisingly highly correlated (0.836), which implies that secondary schools that make more progress than the average secondary school in Mathematics also make more progress in Language than average, and vice versa. Thirdly, the residual variance associated with being a male pupil in Mathematics and the residual variance associated with being a male in Language are not as highly correlated (0.647) as the previous examples, but the pattern is clear: secondary schools that are make a better contribution for male pupils' progress in Mathematics also make better a contribution for them in Language, and vice versa (see: Figure 5.4). Some of these relationships can be appreciated more clearly in Figures 5.3 and 5.4 below.

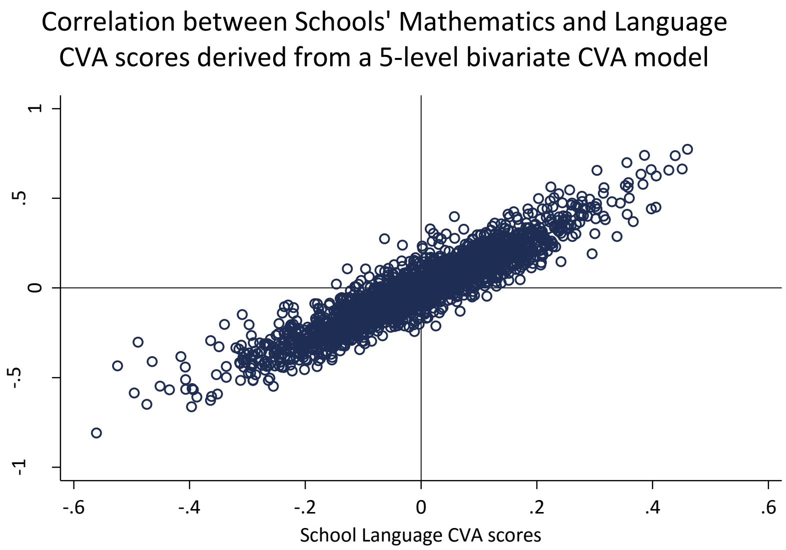


Figure 5.3: Correlation between the random intercepts of Mathematics and Language at the secondary school level

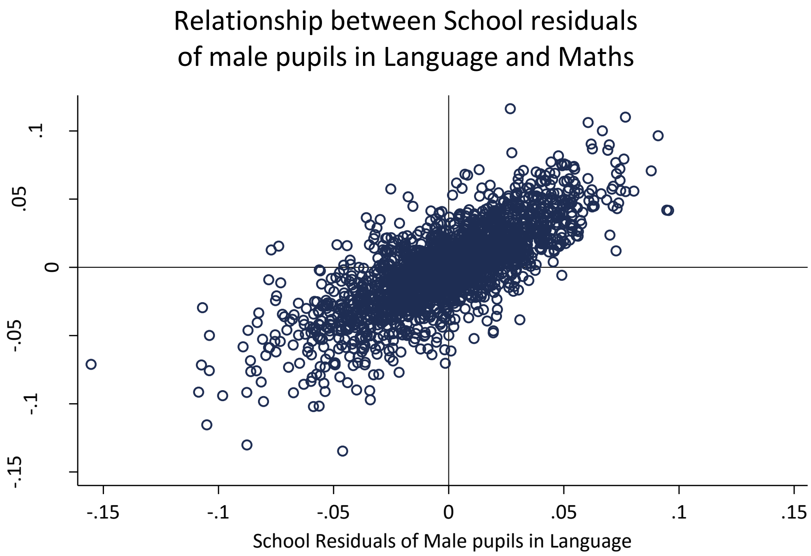


Figure 5.4: Relationship between the residual variance associated with being a male pupil in Mathematics and Language at the secondary school level

On another front, relationships between the intercepts of both outcomes at the other levels of the structure closely resemble those previously reported. At the level of classes, local authorities and primary schools, the correlations between Mathematics and Language are very high (0.784, 0.858 and 0.765, respectively), which means that the averages at all levels in Mathematics and Language are positively correlated. At the level of pupils, as noted above, the relationship between performance in Mathematics and performance in Language reflects on a totally different process than what occurs at the higher levels of the structure. The relatively weak correlation between the two subjects at this level results from the high heterogeneity of pupils and their diverse characteristics, as well as abilities, while at the level of classes, schools (primary and secondary) and local authorities, correlations between the subjects reflect upon the diverse policies, settings, resources, etc.

Table 5.17: Model fit comparison between random coefficients CVA bivariate model and full model (including cross-level interaction effects)

Parameter†	Random coefficients for prior attainment and gender (m4.2)	Full bivariate CVA model (m5)
DIC	613,358.3	613,323.3
pD	13,339.34	13267.52
DIC (m4.2) - DIC (m5)	--	35
Units: Pupils	180,575	180,575
Units: Classes	7,459	7,459
Units: Primary schools	5,537	5,537
Units: Secondary schools	2,438	2,438
Units: Local Authorities	320	320

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: m4.2: 5,000 iterations; m5: 215,000 iterations. In both models: burn-in: 500; thinning: 1.

Table 5.17 shows that the full model significantly improves the overall fit. Although the deviance is not reduced as drastically as with previous intermediate models, the specification of the full bivariate CVA model also reduces the number of effective parameters from the previous model.

In the next section, the discussion is set out to underpin the practical knowledge that can be gained from these results. More specifically, school accountability measures are derived and described in detail.

5.10. What are the implications to school accountability measures derived from this bivariate CVA model?

In a CVA model, a key assumption is that after controlling for all relevant factors that are beyond the control of schools, i.e. pupils' characteristics or compositional effects, as well as non-malleable school characteristics, the remainder of the variance at the level of schools can be used to estimate the amount of value that schools add to their pupils' educational trajectories.

Extending formulae from Chapter 3 to estimate school residuals with more than two levels of variation, CVA scores have been estimated for Mathematics and Language. Figure 5.5 depicts the ranked estimated school CVA scores in both subjects.

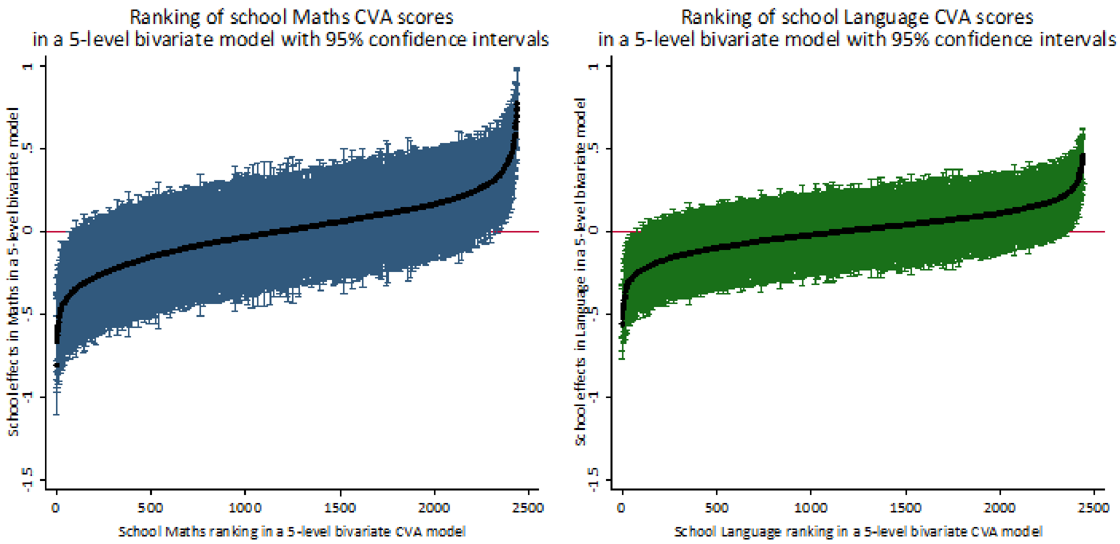


Figure 5.5: Secondary school rankings in Mathematics and Language derived from the 5-level bivariate CVA model

When comparing both sets of estimated CVA scores, it is appreciated that the distribution of school effects in Mathematics is wider than in Language. Furthermore, about 9% of schools are below the national average (red line at 0 in the y-axis) in Mathematics, while about 9% is above average. These percentages in Language are approximately 7%. From these estimated effects, it can be argued that secondary schools add more value in Mathematics than they add in Language to their pupils.

On another front, when comparing residuals from the univariate 4-level models for Mathematics and Language, only a 4% (approx.) are below (and above) average in both subjects. In contrast, when comparing the residuals from the bivariate CVA model for both subjects, there is a higher rate of agreement between subjects. Approximately a 6% of schools are below (and above) the national average.

As seen in Chapter 4, based on whether the confidence interval of a school CVA estimate overlaps or not with the national average, a simple 3-level classification can be derived: 1) schools below the national average; 2) average schools; and 3) schools above the national average. In the bivariate case, though, there are two sets of school effects, i.e. one for Mathematics and another for Language. Therefore, schools can have diverse classifications according to subjects. When schools are classified as above average in one subject but not the other, they are classified as average. The reverse works in the same way: when schools are classified as below average in one subject but not the other, they are classified as average. It is worth noting that no schools are classified as above average in one subject and below average in another. Consequently, there are no huge discrepancies. In Table 5.18, the classifications derived from the traditional CVA model (2-level model) and the extended bivariate CVA model (5-level model) are presented for comparison.

5.18: Comparison of school classifications in the traditional 2-level CVA model and the bivariate 5-level CVA model

	Bivariate CVA model				
Traditional CVA model	**Below average**	**Average**	**Above average**	**Total**	**Percentage‡**
Below Average	130	106	0	236	9.68%
Average	16	1,880	8	1,904	78.10%
Above average	0	162	136	298	12.22%
Total	146	2,148	144	2,438	100%
Percentage†	5.99%	88.11%	5.91%	100%	

Note: The diagonal shows the agreement between the models. 2,146 (88.02%) schools remain in the same category in both models.
† Within bivariate CVA model classifications.
‡ Within traditional CVA model classifications. Classifications were derived from two univariate CVA models.

In Table 5.18 above, it is appreciated that the agreement between both models is fairly high (88.02%). However, discrepancies can be deemed as highly relevant from a substantive point of view. In a traditional 2-level CVA model, 162 (6.64%) schools are wrongfully classified as "above average", whereas they are "average" according to the 5-level bivariate model. In turn, in a traditional 2-level CVA model, 106 (4.35%) schools are unfairly classified as "below average", whereas they are "average" according to the 5-level bivariate model. As was the case in chapter 4, this comparison between models can derive into a two-sided argument. On the one hand, one can argue that schools classified as above average in a traditional 2-level CVA model are unfairly downgraded to average in a 5-level bivariate CVA model, while schools classified as below average in the 2-level model are groundlessly upgraded to average in the 5-level model. On the other hand, one can make the same case, *mutatis mutandis*, to favour the 5-level model. To sort out these allegations, one should bear in mind the purpose of the CVA model. Given that the 5-level model effectively controls for statistically significant sources of variation external to the schools and the pupils that the traditional models do not take into account, one should presume that comparisons derived from it are fairer.

These discrepancies are of the utmost importance when considering the repercussions of schools being classified in certain categories. As discussed previously, in a high-stakes accountability system, such as the one to be implemented in the Chilean education system, underperforming schools can be subject to extreme measures, such as closure. Naturally, the debate is still open with respect to how models for school accountability should be implemented. However, it is apparent from these analyses that the traditional approach is insufficient to deal with the complexity of the school performance phenomenon.

5.11. Conclusions

A 5-level cross-classified bivariate CVA model for progress in Mathematics and Language was implemented to analyse the intricacies of school performance and to draw relevant conclusions about the effectiveness of Chilean schools. The full model is a highly complex specification that attempts to give more insight into how much schools contribute to the educational trajectories of their pupils, by controlling for the most relevant factors that go beyond what any given school can regulate or intervene. This is done with the purpose of isolating the "true" school effects.

From this analysis, it can be concluded, firstly, that progress in Mathematics and Language are undoubtedly related to each other and they need to be analysed accordingly. This is also complementary to the idea that school effectiveness is not a unidimensional phenomenon, because neither schools teach curriculum subjects completely separated from each other nor do pupils learn in an isolated way. Provided that the data are available, school value-added models should take into account all possible relationships between academic outcomes in different subjects, as well as non-academic educational outcomes. A thorough analysis of a model such as the latter, where academic and non-academic outcomes are analysed simultaneously is developed by Timmermans (2012). For this research, there are no data available to analyse neither further subjects nor non-academic outcomes, which is undoubtedly a shortcoming. However, this does not undermine the value of this analysis on its own right, because it brings together a more integrative vision of school effectiveness, along with the necessary technical sophistication.

The second main conclusion from this chapter is that the bivariate CVA model is worthwhile in spite of its ever-increasing complexity. This is due to its power to discriminate better amongst underperforming, outperforming and average schools in a way that is still consistent with less complex models, such as the one presented in the previous chapters, as well as the traditional 2-level CVA model. It is consistent because it successfully identifies broadly the same most and least effective schools as less complex models, while it discards others that were misclassified under those categories, when instead they were actually average-performing.

Thirdly, in this chapter it has also been demonstrated that the analysis of pupils' educational outcomes needs to incorporate information about previously attended schools to estimate the contribution of the currently attended school more precisely. It has been demonstrated that carry-over effects from primary school into secondary school are relevant and their estimation can grant more reliability to school comparisons.

Finally, as mentioned in the previous chapter, this further extension to the traditional CVA models (and the models presented in Chapter 4) embodies a richer and broader concept of school effectiveness. This, in turn, lays out a set of non-trivial and non-negligible adjustments of the utmost relevance when feeding back diverse stakeholders, namely: parents, head teachers, local education authorities, policy makers, etc. This bivariate CVA model is ultimately a valuable tool that takes a more thorough account of the complexity of the school performance phenomenon.

In the next chapter, the analysis turns to a different aspect of academic performance, shifting the focus towards the investigation of the effects of malleable factors with the purpose of informing internal school accountability.

Chapter 6: Exploring the effect of cultural capital on pupils' academic performance for internal school accountability

6.1. Introduction

In Chapters 4 and 5, the focus of the analyses has been set upon the multiple challenges that estimating school value-added entails. It was demonstrated that using traditional CVA models, i.e. a 2-level model for current test scores controlling for prior attainment and socio-economic status at the level of pupils and schools, fails to capture the complexities of academic performance in a way that is not negligible because of its potential consequences when reporting back to stakeholders. In other words, an educational accountability system with incomplete information reported by a less-than-satisfactory statistical model is fundamentally flawed and remedial actions need to be sought with regard to the way in which the data are analysed.

As the aforementioned implies, the previous chapters have been mainly focusing on the variation between secondary schools and how to estimate school effects more precisely. In this chapter, the main focus of the analysis is shifted back to the pupils, while maintaining a careful lookout on the relevant issues around the necessary extensions to the traditional CVA models discussed in the previous chapters.

In this chapter, the analysis mainly aims to illustrate an alternative approach to unravel the complexity of the school performance phenomenon and to ascertain the socio-economic gradient in a way that is more accessible from a public policy viewpoint. This ultimately contributes to inform internal school accountability with the purpose of school improvement. Since previous chapters have been focused on the issue of informing the public in general, parents and/or guardians, stakeholders and government officers about the effectiveness of schools, the most challenging aspects of the analyses that were taken into account are related to methods used and the selection of variables to distinguish the school-level variation more precisely to be able to pinpoint the effects particularly attributable to each school more reliably. This chapter seeks, therefore, to analyse the effect of cultural capital as a meaningful influential factor on pupils' progress over which schools can actually intervene to improve their own practices and results.

In the first section of this chapter, the way in which the cultural capital indicator was constructed using confirmatory factor analysis for ordered categorical data is described. In the next section, the implementation of a series of multilevel models to analyse the effect of cultural capital on Mathematics and Language performance is detailed following the analytical strategy outlined in Chapter 3. Here, many different hypotheses are tested by developing several different specifications of a bivariate multilevel model. Finally, in the last section, some relevant conclusions are drawn from these analyses.

6.2. A measurement model for cultural capital as a latent variable

In order to avoid possible problems of multicollinearity by adding different variables theoretically related to cultural capital and to each other, a latent variable approach can be adopted. As mentioned in Chapter 3, four variables imported from the parents' survey can be thought of as a realisation of the latent construct "cultural capital" (Bourdieu, 1986). These four variables are: a) parents' qualifications; b) number of books in the household; c) frequency with which pupils read "for pleasure" (fiction books, novels and/poetry, magazines, newspapers, etc.); and d) frequency with which parents read "for pleasure".

The choice of items is rooted in the literature as discussed in chapter 2; although it is also constrained by data availability. Additional items related to ownership of other resources such as calculators and computers were considered; however, they were very likely to constitute proxies for (or were very closely related to) family income. These variables were therefore disregarded insofar as they were judged to be redundant in the subsequent multilevel models, in which income was accounted for.

The confirmatory factor analysis model theorises the latent variable "cultural capital" to be caused by the realisation of the four variables already mentioned. The estimated coefficients of this model are displayed in Table 6.1.

The first part of Table 6.1 (below) displays the coefficients (or loadings) for each variable of the measurement model. It is appreciated that the two most important variables in the model are number of books (0.798) and parents' qualifications (0.669). Reading frequencies for pupils and parents are also significant; however their loadings are relatively smaller (0.421 and 0.418, respectively). Meanwhile, the second part of Table 6.1 presents the estimated thresholds for the categories within each variable of the measurement model. Thresholds indicate the point in the theoretical standard normal distribution (mean of 0 and standard deviation of 1) of the latent variable of cultural capital, in which the current category is more likely than the previous one. Given that the categories are ordered in nature, a distinct upwards trend is observed in all thresholds and, as expected, the lowest categories are associated with negative values of cultural capital.

Table 6.1: Estimated coefficients of the measurement model of the cultural capital latent variable indicator

Manifest variables	Coefficients*	Estimate	Std. Err.
Cultural Capital BY	Parents' qualifications	0.669	0.003
	Number of books	0.798	0.003
	Pupils' reading frequency	0.421	0.003
	Parents' reading frequency	0.418	0.003
Thresholds*	**Categories**	**Estimate**	**Std. Err.**
Number of books†	Between 1 and 10 books	-2.256	0.008
	Between 11 and 50 books	-0.771	0.003
	Between 51 and 100 books	0.34	0.003
	More than 100 books	0.914	0.003
Parents' qualifications†	Secondary school	-0.884	0.003
	Vocational education	0.437	0.003
	University degree or postgraduate	0.867	0.003
Pupils' reading frequency†	Never	-1.139	0.004
	Rarely	-0.827	0.004
	Sometimes	0.274	0.003
	Always	1.156	0.004
Parents' reading frequency†	Sometimes every month	-1.335	0.004
	Sometimes every week	-0.364	0.003
	Daily	0.71	0.003

* All estimated coefficients and thresholds are significant at the 0.001 level.
† Reference categories are: None; Primary or less; Pupil does not have any; Never.

In Table 6.2, the measures of goodness of fit of the estimated model are presented. There is abundant literature discussing the diverse goodness of fit measures to be taken into account when specifying latent variable models. Specifically, from the Structural Equation Modelling literature, criteria for judging the goodness of fit of this measurement model were taken from diverse sources Hooper et al. (2008), Yu (2002), Kaplan (2009), Kline (2011) as well as Finney and DiStefano (2006). Firstly, the Chi-square measure does not support the hypothesis that this measurement model is indeed a good model; however, researchers have shown that large sample sizes tend to make this index oversensitive (Kaplan, 2009), and hence alternative measures need to be taken into consideration.

Table 6.2: Goodness of fit of the confirmatory factor analysis for ordered categorical data of the latent variable "cultural capital"

Goodness of fit indices	Value
Chi-square (df, p)	1697.38 (2, p<0.001)
WRMR (Weighted Root Mean Square Residual)	7.526
TLI (Tucker-Lewis Index)	**0.96**
CFI (Comparative Fit Index)	**0.987**
RMSEA (Root Mean Square Error of Approximation)	**0.065**

The WRMR is the modified version of the Standardised Root Mean Square Residual (SRMR), when the WLSMV estimator is used. This index seems to be showing that the model is a poor fit, because its value is far above the recommended value of 1.00 (Finney & DiStefano, 2006; Yu, 2002); however, this index has not been proved to behave stably in simulation studies, overrejecting trivially misspecified models (Yu, 2002) and for that reason, caution is suggested when using this index to assess goodness of fit.

The comparative fit indices to a baseline model (TLI and CFI) indicate a more than acceptable fit, where some researchers recommend to judge a model as acceptable with values over 0.95 for both indices (Hooper et al., 2008; Kaplan, 2009; Kline, 2011). The RMSEA index also shows a fairly good fit with a value of 0.65. Some researchers favour models with RMSEA values below 0.8 (Hooper et al., 2008), but others use 0.5 as the cut off value for RMSEA (Finney & DiStefano, 2006; Kaplan, 2009; Kline, 2011). However, there is no consensus on the threshold for RMSEA, which is why these indices need to be considered with caution, because other authors consider a value around 0.06 to be well-behaved for hypothesis testing under certain circumstances (Yu, 2002).

Given the discussed measures, the measurement model for the cultural capital indicator is judged to have sufficient validity and appropriate fit, as well as pragmatic intrinsic value in relation to the underlying theory. As mentioned before, the estimated latent variable is constrained to have a mean of zero and a variance of one, and also all four measures of the latent variable were allowed to be freely estimated. This is done with the purpose of obtaining standardised factor scores from this model. These predicted standardised factor scores were imported back to Stata (StataCorp, 2011) to run the multilevel models using the user-written module "runmlwin" (Leckie & Charlton, 2013).

6.3. A multilevel analysis of the effect of cultural capital on academic progress

As mentioned previously, the multilevel models implemented in this chapter have been fitted using the MLwiN software (Rasbash, Charlton, et al., 2012) via Stata (StataCorp, 2011) with the "runmlwin" module (Leckie & Charlton, 2013). The estimation method used is the Iterative Generalised Least Squares (IGLS) algorithm and the analytical strategy follows a bottom-up pattern as described in Chapter 3. The models fitted throughout this chapter follow the multilevel analytical strategy described in Chapter 3 and they are briefly described in Table 6.3.

Table 6.3: Models implemented in Chapter 6

Model	Description
Model 0.0	Empty bivariate 2-level model; pupils nested within secondary schools.
Model 0.1	Empty bivariate 3-level model; pupils nested within secondary schools within local authorities.
Model 1	Bivariate 3-level model controlling for prior attainment only (not presented).
Model 2	Bivariate 3-level model controlling for prior attainment and cultural capital.
Model 3.1	Bivariate 3-level model controlling for prior attainment, cultural capital, income and the interaction between cultural capital and income.
Model 3.2	Bivariate 3-level model controlling for prior attainment, cultural capital and income.
Model 4	Bivariate 3-level model controlling for prior attainment, cultural capital, income and average school-level cultural capital.
Model 5	Bivariate 3-level model controlling for prior attainment, cultural capital, income, average school-level cultural capital and the random effects of prior attainment and cultural capital at the school level.

In sum, the multilevel models in this chapter are fitted with the purpose of assessing the impact of cultural capital on pupils' performance in Mathematics and Language. The implemented models are bivariate multilevel models with standardised scores in Mathematics and Language (SIMCE 2006) as the dependent variables, and the standardised cultural capital indicator, alongside its polynomial terms (squared and cubed), as the independent variables. This model also controls for the fixed effects of prior attainment (standardised scores SIMCE 2004) and its polynomial terms (squared and cubed), as well as the random effects of secondary schools and local authorities. The final model also controls for socio-economic and demographic characteristics of the pupils and the schools.

6.3.1. Bivariate variance components and raw value-added models

In this section, the purpose is to set the baseline models of progress as seen in Chapters 4 and 5. The random effects from the levels of classrooms and primary schools have been excluded from this analysis for a number of empirical and theoretical reasons, as well as for parsimony. This is also done with respect to the aim of this chapter, which is to analyse more deeply pupils' heterogeneity and the socio-economic gradient that were described in Chapters 4 and 5.

Firstly, it is empirically plausible that the socio-economic segregation of primary schools does not differ significantly from the segregation found in secondary schools, and as found in the previous chapters, the random effect of primary schools is rather marginal compared to the effect of secondary schools. Secondly, it appears that classrooms have little information to contribute to this analysis, since socio-economic segregation is mainly found between schools, not within schools, which tend to be rather homogenous in terms of SES. Since the model is set up nearly identically as in Chapter 5, the estimated parameters from the variance components model only vary slightly. Results are presented below in Table 6.4.

As expected, the variance components reported in Table 6.4 closely match those reported in Chapter 5. Needless to say, this bivariate variance components model is significantly better fitting than a variance components with fewer levels (2-level model or single-level model) or constrained to have a zero covariance between the two subjects (univariate multilevel models), as also reported in Chapters 4 and 5. Likewise, as seen in previous chapters, a great proportion of the total variance in Mathematics and Language scores is due to factors external to pupils' abilities or characteristics.

Table 6.4: Bivariate variance components models with 2 and 3 levels.

Parameters†	**Model 0.0 (2-level model)**		**Model 0.1 (3-level model)**	
Fixed Part	**Estimate**	**Std. Err.**	**Estimate**	**Std. Err.**
Intercept Maths	0.018	0.015	-0.124	0.024
Intercept Language	0.024	0.013	-0.109	0.022
Random Part	**Estimate**	**Std. Err.**	**Estimate**	**Std. Err.**
Level 1: Pupils				
Variance (Int. Maths)	0.569	0.002	0.569	0.002
Covariance (Int. Maths, Int. Language)	0.347	0.002	0.347	0.002
Variance (Int. Language)	0.670	0.002	0.670	0.002
Level 2: Secondary schools				
Variance (Int. Maths)	0.509	0.015	0.437	0.014
Covariance (Int. Maths, Int. Language)	0.433	0.013	0.367	0.012
Variance (Int. Language)	0.392	0.012	0.331	0.010
Level 3: Local Authorities				
Variance (Int. Maths)	--	--	0.069	0.012
Covariance (Int. Maths, Int. Language)	--	--	0.063	0.011
Variance (Int. Language)	--	--	0.058	0.010
Model fit information				
Deviance	888,795.38		888,601.25	
Chi-squared (df, p)	--		194.13 (3, p<0.001)	
Number of parameters	8		11	
AIC	888,811.38		888,623.25	
N	202,605		202,605	

† Obtained via IGLS estimation.

These models do not control for prior attainment and hence are not true value-added models, as defined by Goldstein (1997). The next step in the modelling process is to set up a model that controls for these measures in both outcomes; however, this step is not presented here since it does not provide further information for the purpose of this chapter. This raw bivariate model (model 1) closely matches the one reported in Chapter 5, with the exception of having fewer levels.

6.3.2. The non-linear effect of cultural capital on Mathematics and Language performance

After specifying the raw bivariate VA models, the cultural capital indicator described in the first section was put into the model. As mentioned earlier, "cultural capital" is a latent variable, which implies that it is a relative measure whose distribution is imposed and its estimation is heavily reliant upon the available data. This means that even though this estimated measure is intended to be used as an explanatory variable in the model, its "true" effect is impossible to observe. In any case, the use of the latent variable approach allows capturing information from diverse sources, removing unwanted disturbances in the data and estimating the effect of a conjuncture of variables, which seems to be appropriate to analyse the effect of a rather elusive concept, such as cultural capital.

Table 6.5: Bivariate 3-level model controlling for prior attainment and cultural capital (model 2)

Fixed part	Parameters†	Coef.	Std. Err.	95% C.I.	
Mathematics	Intercept	-0.057	0.010	-0.077	-0.036
	Prior attainment	0.635	0.002	0.630	0.640
	Prior attainment squared	0.035	0.001	0.033	0.037
	Prior attainment cubed	-0.022	0.001	-0.024	-0.021
	Cultural capital	0.076	0.003	0.070	0.082
	Cultural capital squared	0.001	0.002	-0.002	0.005
	Cultural capital cubed	-0.003	0.001	-0.006	-0.001
Language	Intercept	-0.054	0.008	-0.070	-0.038
	Prior attainment	0.667	0.003	0.662	0.672
	Prior attainment squared	0.026	0.001	0.023	0.028
	Prior attainment cubed	-0.032	0.001	-0.033	-0.030
	Cultural capital	0.093	0.003	0.086	0.100
	Cultural capital squared	0.013	0.002	0.010	0.017
	Cultural capital cubed	0.004	0.002	0.001	0.007
Random part	**Parameters†**	**Coef.**	**Std. Err.**	**95% C.I.**	
Level 3: Local Authorities	Variance (Int. Maths)	0.008	0.002	0.004	0.012
	Covariance (Int. Maths, Int. Language)	0.006	0.002	0.003	0.009
	Variance (Int. Language)	0.005	0.001	0.002	0.007
Level 2: Secondary schools	Variance (Int. Maths)	0.114	0.004	0.107	0.121
	Covariance (Int. Maths, Int. Language)	0.077	0.003	0.072	0.083
	Variance (Int. Language)	0.066	0.002	0.061	0.070
Level 1: Pupils	Variance (Int. Maths)	0.314	0.001	0.312	0.316
	Covariance (Int. Maths, Int. Language)	0.098	0.001	0.096	0.099
	Variance (Int. Language)	0.387	0.001	0.385	0.390
Model fit					
Deviance	697,875				
Number of	23				
AIC	697,921				
N	199,061				

† Obtained via IGLS estimation.

‡ Chi-square test is not reported here, because no model comparison is intended. However, one can use the information reported to derive such comparisons.

The model presented in Table 6.5 is set up to specify the linear effect of prior attainment as well as cultural capital alongside its non-linear effects; this means that cultural capital has been squared to fit a curve to control for "floor effects" and cubed to fit a curve to control for "ceiling effects". Alongside these non-linear effects of cultural capital, non-linearity is also suspect and proved to be significant in the case of prior attainment in both subjects; these polynomials are also included to differentiate between bottom and ceiling effects due to cultural capital and due to prior attainment.

Polynomial terms are somewhat hard to interpret, therefore plotting the predictions from this model can be useful. Below, Figure 6.1 depicts these relationships.

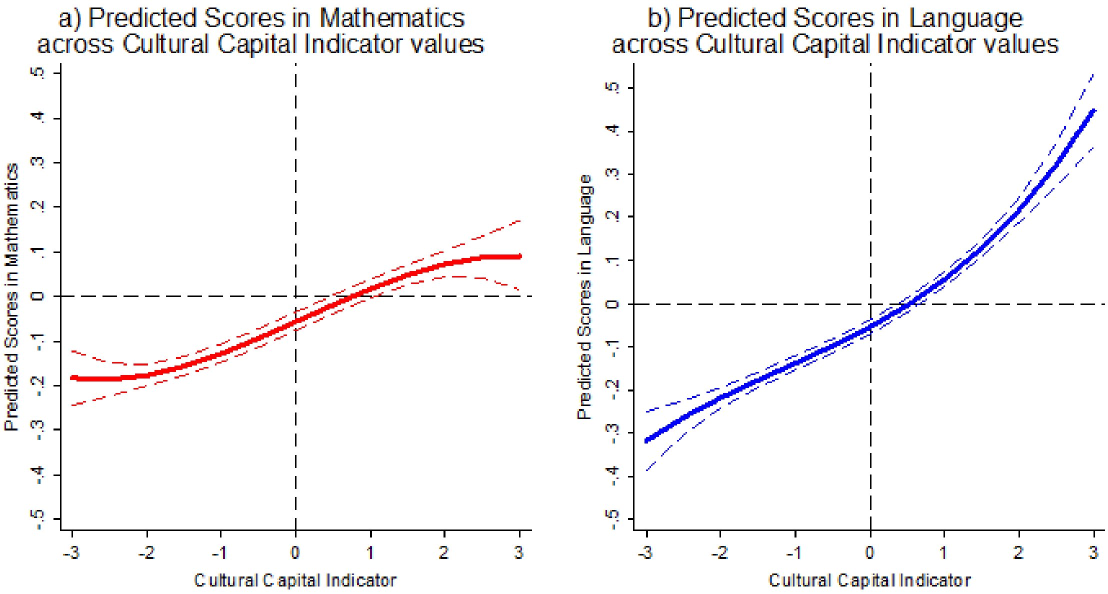


Figure 6.1: Predicted standardised SIMCE scores from bivariate 3-level model controlling for prior attainment, cultural capital and the random effects of secondary school and local authority.

In Figure 6.1 a), it is appreciated that between -2 and 2 standard deviations of cultural capital the relationship between Mathematics scores and cultural capital tends to follow a straight pattern, although the line is not very steep. On another front, at extremely low values of cultural capital (less than -2 standard deviations), the predicted trajectory of progress in Mathematics is flat, which means that pupils with low cultural capital need to overcome a greater gap to make progress in Mathematics than pupils with higher cultural capital. On the other hand, at extremely high values of cultural capital (more than 2 standard deviations), the predicted line tends to flatten once more. This decrease in the growth rate of progress in Mathematics indicates that pupils with extremely high cultural capital can only make limited progress.

In Figure 6.1 b), a rather distinct pattern is observed, especially at the extremes of the distribution of cultural capital (x-axis). An ascending curve is observed at the lowest values (less than -2 standard deviations) of cultural capital, which implies that little increases in cultural capital can make a larger difference than further increases at values between -2 and 2 of cultural capital. Then, between -2 and 2 standard deviations of cultural capital, a somewhat straight line is recorded that is steeper than the straight line in Mathematics attainment, which implies that the relationship between cultural capital and Language attainment is stronger. Finally, at the highest values of cultural capital (beyond 2 standard deviations), another ascending curve is recorded, which implies that pupils with extremely high levels of cultural capital have an ever steeper rate of progress than any other pupil.

6.3.3. The effect of cultural capital versus the effect of income on academic performance

Naturally, when analysing the effect of cultural capital, the existence of an underlying (perhaps even confounding) effect on academic performance caused by income becomes suspect. By specifying income as yet another explanatory variable in the model, it can be hypothesised that the effect of cultural capital will tend to decrease or disappear altogether. Furthermore, given that a relationship between cultural capital and income is indeed plausible, it is also plausible that the effect of cultural capital on pupils' scores varies by income intervals.

The following models explore these hypotheses by specifying the set of dummy variables related to average monthly household income (as reported in the parents' SIMCE survey) as explanatory variables, with low income as the reference category, as well as the interaction term between the cultural capital index and each of the income categories (again with low income as the reference).

In Table 6.6, it is observed that none of the dummy variables that account for the interaction between income and cultural capital turned out to be significant in the case of Mathematics and only one of them in the case of Language, i.e. the interaction between high income and cultural capital, where a significant positive difference is found with respect to low income children. Nevertheless, since this is a set of variables affecting the outcome as a whole, their significance needs to be analysed accordingly via a joint Wald test, which failed to demonstrate significance. Moreover, when comparing the fit of models 3.1 and 3.2, it is apparent that by removing the interaction term between income and cultural capital, the overall fit is improved, which is revealed by the lower AIC value of model 3.2. The most parsimonious model is therefore 3.2, where no interactions are specified.

Table 6.6: Model comparison between bivariate model with (m3.1) and without (m3.2) a control for the interaction between income and cultural capital

Fixed part†	Mathematics		Language	
	Model 3.1	**Model 3.2**	**Model 3.1**	**Model 3.2**
	Coef. (s.e.)	**Coef. (s.e.)**	**Coef. (s.e.)**	**Coef. (s.e.)**
Intercept	-0.089 (0.011)***	-0.089 (0.011)***	-0.046 (0.008)***	-0.046 (0.008)***
Prior att.	0.63 (0.002)***	0.63 (0.002)***	0.665 (0.003)***	0.665 (0.003)***
Prior att. squared	0.034 (0.001)***	0.034 (0.001)***	0.026 (0.001)***	0.026 (0.001)***
Prior att. cubed	-0.022 (0.001)***	-0.022 (0.001)***	-0.031 (0.001)***	-0.031 (0.001)***
Cultural capital	0.077 (0.004)***	0.075 (0.003)***	0.089 (0.004)***	0.09 (0.004)***
Cult. capital-squared	0.001 (0.002)	0.0001 (0.002)	0.009 (0.003)***	0.012 (0.002)***
Cult. capital-cubed	-0.003 (0.002)*	-0.003 (0.001)*	0.002 (0.002)	0.004 (0.002)*
Male	0.062 (0.003)***	0.062 (0.003)***	-0.042 (0.003)***	-0.042 (0.003)***
Low-mid income	0.008 (0.003)*	0.008 (0.003)*	0.019 (0.004)***	0.018 (0.004)***
Up-mid income	0.02 (0.007)**	0.016 (0.005)**	0.032 (0.007)***	0.035 (0.006)***
High income	0.028 (0.012)*	0.039 (0.008)***	0.031 (0.013)*	0.053 (0.008)***
Low-mid inc. & Cult. Cap.‡	-0.004 (0.005)		0.002 (0.006)	
Up-mid inc. & Cult. Cap.‡	-0.01 (0.008)		0.009 (0.009)	
High inc. & Cult. Cap.‡	0.009 (0.011)		0.029 (0.012)*	
Model fit information	**Model 3.1**	**Model 3.2**		
Deviance	643,102.44	643,111.38		
Number of parameters	37	31		
AIC	643,176.44	643,173.38		
N	183,142	183,142		

* p<0.05; ** p<0.01; *** p<0.001.

† Obtained via IGLS estimation. Random-effects parameters are not included here for presentational purposes.

‡ Joint Wald test is not significant. Chi-squared=8.92 on 6df, p=0.1781.

From this analysis, it can be concluded that neither income moderates the relationship between cultural capital and progress nor cultural capital moderates the relationship between income and progress, which is the same to assert that the effect of cultural capital does not vary according to income or vice versa. This could also seem to indicate that income and cultural capital may have distinct effects on progress in Mathematics and Language, measuring different aspects of the phenomenon.

After removing the interaction term between income and cultural capital, the model has been simplified to a more parsimonious form. However, the only explanatory variables that are specified so far are at the level of pupils, which implies that no controls have been set up to analyse the influence of the school context. Educational research has demonstrated that peer effects are an important source of variation in school performance, which is why it is highly plausible that there is a contextual component to the effect of cultural capital on performance in Mathematics and Language. This is step 4-onwards of the analytical strategy outlined in Chapter 3.

6.3.4. The contextual effect of cultural capital

So far, the bivariate multilevel models presented in this chapter have only been fitted with pupil-level explanatory variables. However, it is widely known that the school context can influence individual pupils' performance. For instance, a high-achieving (on average) school could boost low-achieving pupils' performance in a way that low-achieving (on average) schools could not. Likewise, cultural capital may operate in the same fashion; schools with high average levels of cultural capital (in comparison to other schools) may foster better performance from pupils with low cultural capital (or lower-than-average in their own schools). In this respect, Torche (2005) refers to the Chilean case where peer effects resulting from school sorting may be a mechanism through which the gap between schools grows. Since this sorting is mainly economically driven, where parents and carers have limited school choice according to their purchasing power, then the school sorting is also related to socio-economic and cultural segregation.

In this section, the effects of the school-level explanatory variables are analysed. Firstly, the set of dummy variables to indicate institutional type (State-funded, Subsidised Independent or Independent) of the secondary school was included as in Chapters 4 and 5. Furthermore, the contextual variable "school average cultural capital", which is a summary of the pupil-level cultural capital index (for parsimony, only the linear term without polynomials) at the secondary school level, was also added. Alongside these two additional higher-level explanatory variables, the following interaction terms were specified: interaction between school type and cultural capital at the pupil level, between school type and school average cultural capital, and between cultural capital at the pupil level and school average cultural capital.

This specification allows to investigate some relevant intermediate research questions: a) Does the effect of cultural capital at the pupil level vary across school types?; b) Does the effect of school type vary across diverse school contexts in terms of average cultural capital?; and c) Does the effect of cultural capital at the pupil level vary across school contexts with diverse average cultural capital?

In Table 6.7, below, several interesting relationships are proved to affect pupils' progress in Mathematics and Language significantly. Given that interaction effects have been specified in this model, the main effects of the fixed part are not interpretable in their own right. When interacting the institutional type of the schools with pupils' cultural capital, it is appreciated that at the same level of cultural capital, pupils in subsidised independent schools and independent schools make significantly less progress than pupils in State-funded schools in

Mathematics. In the case of Language, when comparing pupils with the same level of cultural capital, pupils in state-funded schools make more progress only with respect to pupils in subsidised independent schools, and they do not differ significantly from pupils in independent schools. The estimated random-effects parameters have been suppressed from Table 6.7 for they are not of primary interest in this analysis, although they are relevant controls for the model to work properly. Full details can be found in Appendix 3.1.

Table 6.7: Fixed-effects parameters of the full bivariate model including cross-level interaction effects

		Mathematics	**Language**
	Parameters†	**Coef. (Std. Err.)**	**Coef. (Std. Err.)**
Main effects	Intercept	-0.065 (0.018)***	-0.025 (0.014)
	Prior attainment	0.627 (0.003)***	0.663 (0.003)***
	Prior attainment squared	0.04 (0.001)***	0.029 (0.001)***
	Prior attainment cubed	-0.021 (0.001)***	-0.031 (0.001)***
	Cultural capital	0.076 (0.004)***	0.092 (0.005)***
	Cultural capital-squared	0.009 (0.002)***	0.016 (0.003)***
	Cultural capital-cubed	-0.002 (0.002)	0.004 (0.002)*
	Male	0.062 (0.003)***	-0.041 (0.003)***
	Low-mid income	0.001 (0.003)	0.011 (0.004)**
	Up-mid income	-0.0004 (0.006)	0.018 (0.006)**
	High income	0.008 (0.008)	0.01 (0.009)
	Subsidised Independent	0.024 (0.018)	0.022 (0.014)
	Independent	0.003 (0.077)	-0.141 (0.063)*
	School average Cult.Cap.	0.377 (0.031)***	0.306 (0.024)***
Interaction effects	Subs. Indep. & Cultural Capital	-0.021 (0.005)***	-0.019 (0.005)***
	Indep. & Cultural Capital	-0.03 (0.013)*	-0.02 (0.014)
	Subs. Indep. & School average Cult. Cap.‡	0.063 (0.035)	0.013 (0.027)
	Indep. & School average Cult. Cap.‡	0.094 (0.083)	0.209 (0.068)**
	Cult. Cap. & School average Cult. Cap.§	-0.017 (0.007)*	-0.006 (0.007)
Model fit information	**Deviance**	640,503	
	Number of parameters	54	
	AIC §§	640,611	
	N	183,142	

* p<0.05; ** p<0.01; *** p<0.001.

† Obtained via IGLS estimation. Random-effects parameters are not included here for presentational purposes.

‡ Joint Wald test is significant. Chi-squared=18.55 on 4df, p=0.001.

§ Joint Wald test is significant. Chi-squared=6.47 on 2df, p=0.039.

§§ Overall fit is improved with this specification. AIC from previous model is 643,173.38 (Table 6.6).

When analysing the school-level effect of cultural capital, another interesting feature can be appreciated. The interaction between school institutional type and the school average cultural capital is only significant for the dummy variable corresponding to independent schools and school's cultural capital in the case of Language. This means that the average level of cultural capital in independent schools represents an advantage for their pupils, but only in Language, where they are expected to outperform the rest of pupils. This is unsurprising as independent schools are expected to have a higher average of cultural capital.

Finally, results shown in Table 6.7 also suggest that pupils may significantly benefit from the average cultural capital of the school they attend to, because the higher the school-level average cultural capital, the lower the coefficient of the individual pupil's cultural capital. This implies that, for instance, from two pupils with relatively low cultural capital, the one that is predicted to make the most progress is the one attending the school with the highest cultural capital.

All the aforementioned complex relationships can be more easily observed in graphs as follows. Firstly, the overall relationship between pupils' cultural capital and progress in Mathematics and Language has not changed greatly as can be seen in Figure 6.2. However, some features do stand out (in comparison to Figure 6.1). The most relevant and noticeable difference between Figure 6.2 and Figure 6.1 is that the range of predicted scores has become slightly narrower with predicted standardised scores in Mathematics going from -0.1 to just under 0.15 and from just under -0.2 to just above 0.4 in Language. Predicted scores in Figure 1 went from -0.2 to 0.1 in Mathematics and from -0.3 to 0.4 in Language. This unsurprising range contraction is most certainly the effect of controlling for income, gender, school type and school average cultural capital. Nevertheless, the overall effect of cultural capital prevails confirming the patterns depicted and described in Figure 6.1.

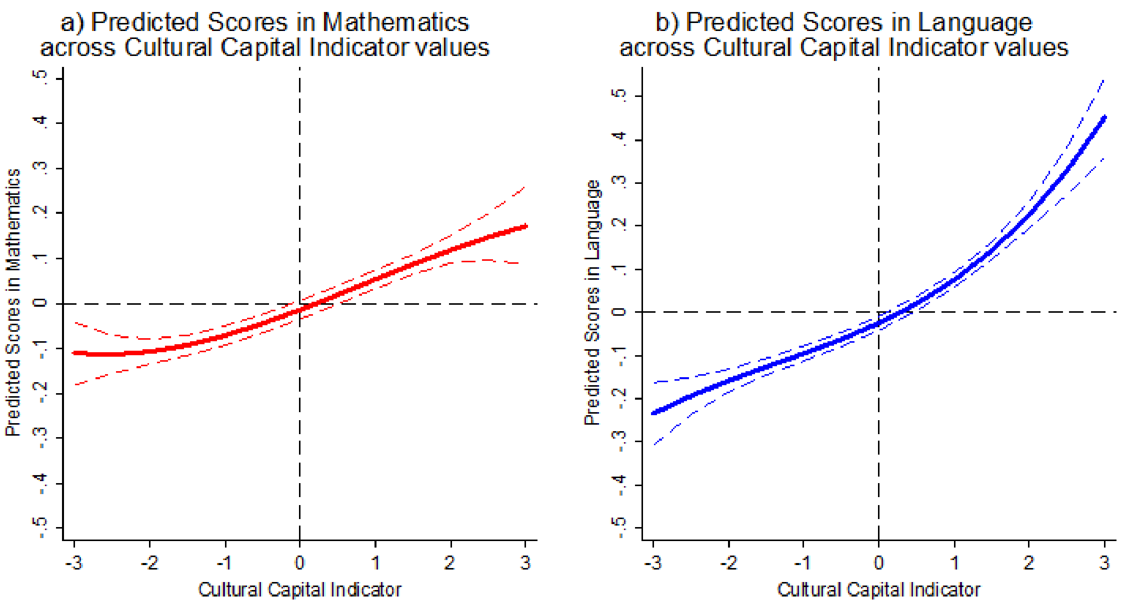


Figure 6.2: Predicted standardised SIMCE scores from full bivariate 3-level model.

The interactions previously presented in Table 6.7 also revealed relevant relationships between the explanatory variables and the progress of pupils in Mathematics and Language. In the following graphs, the effect of cultural capital on academic performance is moderated by the school-level variables institutional type and school average cultural capital.

In Figure 6.3, it is appreciated that the slope of the relationship between cultural capital and the standardised scores in Mathematics is steeper for pupils attending State-funded schools. Meanwhile, the slope of the relationship between cultural capital and the standardised predicted scores in Mathematics is noticeably shallower for pupils in subsidised-independent and independent schools. This implies that cultural capital has the lowest impact on Mathematics performance on pupils attending subsidised independent and independent schools. Conversely, pupils who could take the greatest advantage from increasing their cultural are potentially those who attend State-funded schools.

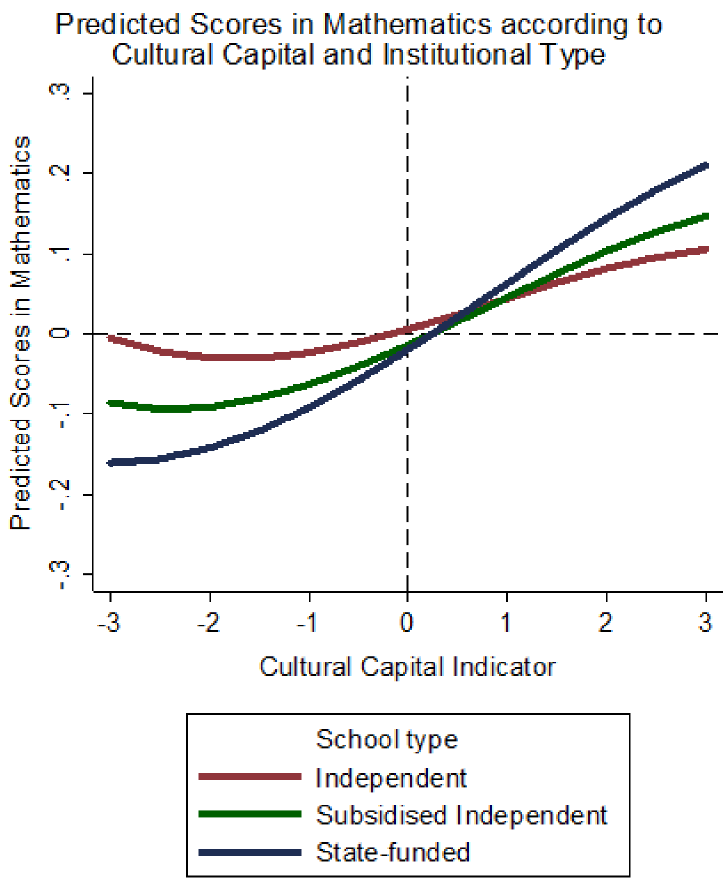


Figure 6.3: Predicted standardised SIMCE scores in Mathematics against cultural capital index scores by institutional type of the school.

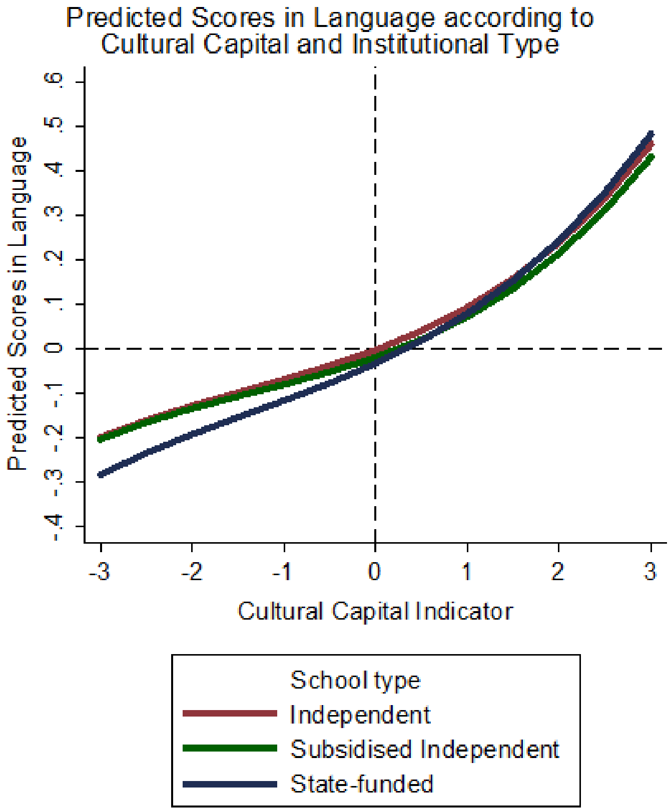


Figure 6.4: Predicted standardised SIMCE scores in Language against cultural capital index scores by institutional type of the school.

As appreciated in Figure 6.4, the shape of these estimated trajectories in the case of Language are quite different. Firstly, schools are not that clearly differentiated in terms of their institutional type as they are in the case of Mathematics. Inspecting the curves of subsidised independent and independent schools, hardly any differences are found. However, the most noticeable feature of this graph is again that pupils in State-funded schools benefit the most from increasing cultural capital. It is particularly remarkable that at the lower end of cultural capital values the curve is steeper, indicating that increasing levels of cultural capital for pupils with very low cultural capital has a greater impact on pupils in State-funded schools compared to pupils in other school types who make a more steady progress in Language.

In summary, from what is depicted in Figure 6.4, it can be observed that increasing individual pupils' cultural capital does not make a major difference when pupils are situated around the average or above in the distribution of the cultural capital. However, it does make a remarkable difference for pupils with extremely low cultural capital who attend State-funded schools.

On another front, Figures 6.5 and 6.6 depict the cross-level interaction between individual pupils' cultural capital and school-level average cultural capital, which has been reported in Table 6.7. In both figures a somewhat similar overall trend is appreciated with only a few peculiarities for each subject. These figures show that the progress that pupils make depends not only on their own cultural capital, but also on the average cultural capital of the school they attend.

In the case of progress in Mathematics (Figure 6.5), when grouping schools according to different mean values of the school-level cultural capital, the differences are more pronounced than in Language and they are nearly all significant across the range of individual cultural capital index scores. This implies that from two pupils with the same level of prior attainment, same demographic and socio-economic characteristics and same level of cultural capital, the pupil that will most likely outperform the other pupil in Mathematics is the one who attends the school with the highest average level of cultural capital. This could be considered as evidence in support of the "peer effects" hypothesis. Nonetheless, this finding is by no means conclusive and other analyses are required to corroborate; for instance, an analysis of the relationships between pupils within secondary schools and how these affect their educational achievement. This is an issue for further research that will be discussed later.

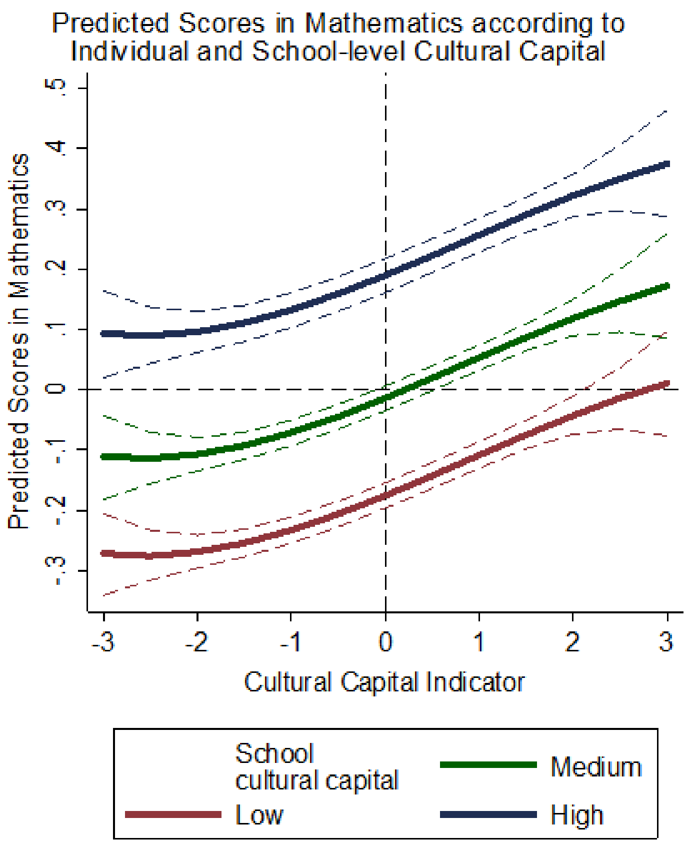


Figure 6.5: Predicted standardised SIMCE scores in Mathematics against cultural capital index scores by school average cultural capital

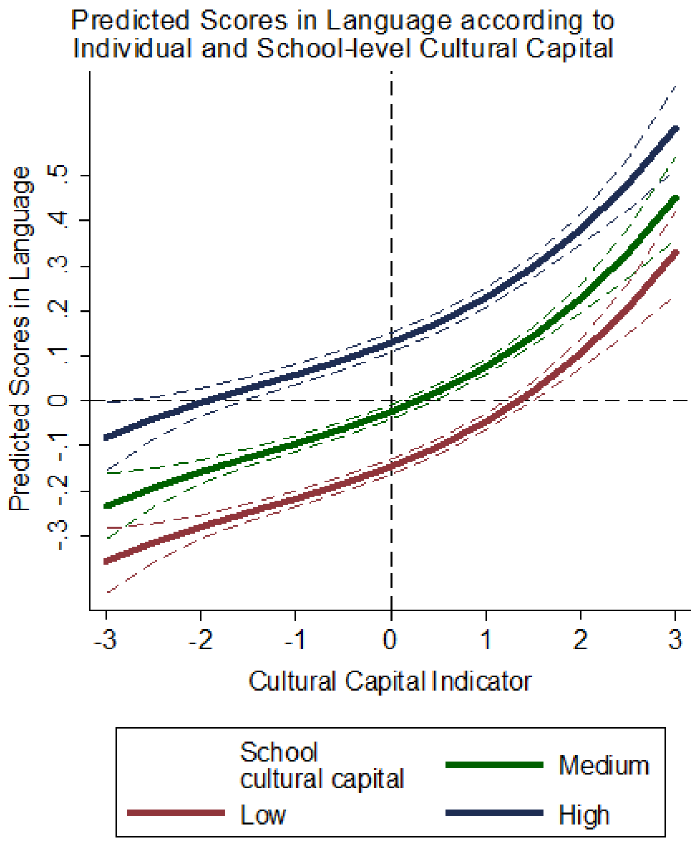


Figure 6.6: Predicted standardised SIMCE scores in Language against cultural capital index scores by school average cultural capital.

Progress in Language (Figure 6.6) follows a slightly different pattern. Here, the differences tend to disappear at the extreme values of individual cultural capital. After 2 standard deviations at both ends of the distribution, the 95% confidence intervals around the trajectory of the relationship between individual cultural capital and progress in Language for different mean values of school-level cultural capital start to overlap. This implies that pupils with extremely low or extremely high cultural capital are not necessarily expected to be significantly better off in any other school, when holding all the other variables constant. Although in a more normal range, this is between -2 and 2 standard deviations, the pattern is very similar to what is found in Mathematics, i.e. when holding everything in the model constant, pupils with most progress are those in schools with higher average levels of cultural capital.

6.4. Conclusions

A bivariate multilevel model for pupils' performance in Mathematics and Language was implemented to analyse multiple hypotheses about the effect of cultural capital on academic outcomes with the ultimate purpose of informing school internal accountability processes. The full model controls for the fixed pupil-level effects of prior attainment and the estimated index of cultural capital alongside their polynomial non-linear terms, as well as gender and average monthly household income. It also controls for the fixed school-level main effects of institutional type and school average cultural capital. Additionally, it controls for the interaction effects between school type and pupils' cultural capital, between school type and school average cultural capital, as well as the interaction between pupils' cultural capital and schools' cultural capital. The full model also controls for the random effects of attending a particular school and living in a particular local authority as well as the random effects or prior attainment at the school level.

An arguable caveat of the models presented in this chapter is that one could assert that, theoretically, scores in standardised tests can also be considered a realisation of cultural capital. This would potentially raise the issue of endogeneity. Nevertheless, the measurement model for cultural capital is constructed with lagged variables, and hence the effect of cultural capital has a temporal dimension. What the models in this chapter show is that the accumulated cultural capital up to year 8 has a sizable and non-negligible effect on the progress that pupils in year 10 make in Language and Mathematics. The value of these models lies precisely in showing that the differences in the processes of accumulation of cultural capital that pupils undergo throughout their schooling and family life, have a subsequent effect on their academic performance.

This model has successfully proved its practical adequacy for analysing the complexities of the effects of cultural capital on educational outcomes. As a result of this specification, several relevant findings could be unveiled. Firstly, cultural capital affects performance in Mathematics and Language differentially; Language is the subject most affected by cultural capital, although the effect on Mathematics is still non-negligible. In both subjects, however, the effect is overall the same: pupils who are culturally advantaged tend to outperform culturally-disadvantaged children, which is consistent with a Bourdieuvian theoretical perspective.

Secondly, cultural capital may also differently affect pupils' performance in combination with other conditions, namely: at the disaggregated pupil level in combination with the institutional school type, at the aggregated school level combined again with school type, as well as combining the effect of cultural capital at the disaggregated pupil level with the aggregated school level. What these interactions show is that the effect of an individual pupil's cultural capital varies according to the characteristics of the school the pupil attends. Furthermore, these variations occur in a way that is consistent yet again with a Bourdieuvian theoretical perspective, because as the average level of cultural capital of a particular school increases, pupils' expected progress increases, which implies that children can take significant advantage of the school's socio-economic and cultural environment. This is also related to third main conclusion drawn from this analysis.

Thirdly, concerning the combined effect of income and cultural capital, the data have failed to demonstrate such an interaction, which makes plausible the hypothesis that their effects have diverse mechanism through which they operate. This is not necessarily against the Bourdieuvian hypothesis that states that socio-economically hegemonic classes sustain their own hegemony through cultural capital. This is because, in the specific case of the Chilean education system, the effect of income can operate also through school choice, which is rather limited to parents' purchasing power. Having noted the foregoing and the fact that pupils from the poorest households generally attend State-funded schools, the significant interaction between cultural capital and school type sheds light over the matter by unveiling that even though State-funded pupils make more progress than pupils in schools of other types, their disadvantaged position is noticed in very low cultural capital contexts. This is because their trajectories start from lower attainment than other pupils, so there is greater room for progress and their disadvantage is even more apparent in terms of current attainment, as opposed to progress, which takes prior attainment into account.

To sum up, these results suggest that cultural capital, materialised in reading habits, access to books and parental qualifications, can make a considerable difference in school performance. In terms of public policy, this has important implications insofar as schools and local education authorities could use this information to take action and potentially boost pupils' academic outcomes by simply implementing policies, for instance, to foster children and parents reading and to seek alliances with publishers to make books more accessible for low-income families. Most importantly, making all kinds of reading materials more accessible to the general population in Chile would be relatively easy to achieve by reducing (or eliminating) the very high taxation on books, which is currently 19%. Nevertheless, this analysis did not intend neither to be exhaustive nor to be reductionist, but to approach a very complex problem with appropriate tools that allow identifying clear action points to be eventually implemented in public policy.

Chapter 7: Discussion

This thesis has analysed the Mathematics and Spanish Language progress of pupils in Chilean schools who moved from primary to secondary school in the years 2004 and 2006. The models implemented in this thesis follow the tradition of school value-added research and extend its reach to give more insight into pupils' progress and the contribution of schools to it. Rather than a conflicting view, this thesis provides further support for previous research; however, it also goes beyond to suit the needs and challenges that the particular national context poses to the analysis.

7.1. The necessity of more complex multilevel models for analysing academic performance and school accountability

Firstly, this thesis underlines the necessity of specifying more reasonably complex CVA models to tackle the complexity of the school performance phenomenon. From a univariate perspective, this implied first and foremost including additional levels of variation to contextualise the knowledge derived from the models even further to what can be achieved by implementing 2-level models. It was seen in Chapter 2 that there are a few studies in which the level of classrooms has been successfully specified and proved to be relevant (Cervini, 2009a, 2009b; Martínez, 2012; Murillo and Roman, 2011). Likewise, the level of local authorities has also been found to be relevant in other studies (Cervini, 2009a, 2009b; Leckie, 2009; Plewis, 2011; Rasbash et al., 2010). Furthermore, the relevance of the specification of the level of primary schools has also been demonstrated by some authors (Leckie, 2009; Goldstein et al., 2007). What these authors stress directly and indirectly is the need for analysing the complexity of academic performance with appropriate and reasonably complex models. As seen in Chapter 4 of this thesis, this is done successfully in terms of the statistical superiority and the valued practical implications of the extended CVA 4-level models with respect to the traditional 2-level CVA models.

Results from the CVA models presented in this thesis reveal the extent of the inequalities of the Chilean education system. Evidence from previous research in the United Kingdom has found that schools only account for between 5% and 20% of the total variation (Rasbash et al., 2010). However, as seen in Chapter 4, a much larger percentage is found in the Chilean education system, where schools account for nearly half the variation in Mathematics tests scores; this is a comparison between 2-level empty models. Furthermore, variation between local authorities reveals that Chilean schools have different levels of performance according to their geographical location. This most likely reflects upon a highly unequal distribution of wealth and resources amongst different areas in Chile. Furthermore, this is certainly an outstanding difference with respect to what has been found in UK-based research, where Local Authorities only account for between 1% and 3% of the total variation (Rasbash et al., 2010). These are signs of the massive differences that can be found in a highly unfair and unequal education system such as the Chilean system. Naturally, this evidence supports the need for more complex analyses. On another front, the extension of the traditional CVA models to adapt to more complexity can also be useful to explore differential effects for certain groups not only at the level of schools (as seen for instance in: Kyriakides, 2004; Plewis, 2011), but also at the levels of classrooms and local authorities. This can be potentially relevant information for educational planning originated at the level of the schools themselves or at the level of the local authorities. In sum, from a general perspective, ascertaining details about how academic performance occurs at different levels for different groups of pupils is crucial not only for a fair accountability system, but more importantly for a fair education system.

7.2. The multidimensionality of academic performance

As a second main point of general discussion, this thesis takes another step forward and engages in the debate about the narrowness of the concept of school effectiveness that the traditional approach entails. As discussed in Chapter 2, analysing the progress of pupils in different subjects in separate models implies that pupils learn subjects disassociated from each other and that schools teach subjects with no reference to each other. This is rather unrealistic from a substantive and a statistical point of view, because connections can exist irrespective of whether this is done intentionally or not at any level. In this thesis, the focus is on progress in Mathematics and Language, which several researchers (Cowan et al., 2005; Donlan et al., 2007; Simmons & Singleton, 2008; Hecht et al., 2001; Vukovic & Lesaux, 2013) have found to be at a certain extent associated with each other at the individual/pupil level. Sulis and Porcu (2014) provided further evidence on the relationship between Mathematics and Language attainment not only at the individual level, but also at multiple levels, more specifically at the level of schools and geographical areas.

This thesis has found strong evidence of varying degrees of association between Mathematics and Language, not only at the level of pupils, but also at the levels of classrooms, primary schools, secondary schools and local authorities. What is also further revealing is that prior attainment follows basically the same patterns: variations in progress, i.e. linear relationship between prior and subsequent attainment, in Mathematics at the level of secondary schools are associated with variations in progress in Language.

Nevertheless, a more thorough analysis of the relationships between subjects would also need to include other subjects as well as non-academic outcomes, such as motivation, satisfaction, etc. Timmermans (2012) carries out such analyses to find that non-academic outcomes are not so strongly associated to academic outcomes as they are to each other; however, these analyses are conducted in the Netherlands, a much more highly egalitarian and developed country with a successful education system, which is rather diametrically opposed to the highly segregated Chilean education system (Valenzuela et al., 2008).

For this thesis, there are no data available to analyse neither additional subjects other than Mathematics and Language nor non-academic outcomes, such as motivation, satisfaction, aspirations, etc., at the pupil-level. Nevertheless, SIMCE reforms in recent years will make possible to analyse further academic subjects which will be contained in subsequent waves of data for several cohorts of pupils at different time points (Years 2, 4, 6, 8 and 10). This is also briefly addressed in Chapter 8. This could be considered as a shortcoming, given that the multidimensionality of school performance has been stressed throughout this thesis; however, this does not undermine the value of these analyses in their own right, because the models presented herein (Chapter 5, more specifically) bring together a more integrative vision of school effectiveness along with the necessary model complexity.

7.3. Issues surrounding external school accountability models

The models implemented in this thesis follow the basic guiding principle that a school accountability system should be, first and foremost, fair to all schools and pupils (See for example: Fitz-Gibbon, 1997; OECD, 2008). As discussed in Chapter 2, this implies that schools should ideally be assessed based only on the practices and circumstances over which they have full control, which requires meticulous work in isolating the school effects from other spurious effects that are favourable to some schools and detrimental to others. Since there is strong evidence of geographical differences, within-school streaming, carry-over effects from primary school and correlation between Mathematics and Spanish Language attainment (as presented in Chapters 4 and 5), holding everything else in the models constant, one should be inclined to regard the classifications derived from the extended CVA models from Chapters 4 and 5 as fairer than the classifications derived from the traditional CVA model. Nevertheless, there are still some issues that may be problematic and/or controversial, which would need further clarifications.

From the perspective of external school accountability, one potential problem of the 4-level CVA models fitted in this thesis is that some would argue that they partially remove effects truly belonging to the schools when specifying the classroom level. This is simply because classrooms are fully nested within schools. Nevertheless, this is a twofold problem and there is no simple answer, because removing the classroom level would result unequivocally in overestimating the pupil-level effects and retaining it could be said to underestimate the school effects. As seen in Chapter 4, when specifying the classroom level, the variance from the pupils and the schools are both reduced, meaning that the classroom effects are a sort of combination between pupil and school effects. As discussed in Chapter 4, this finding is consistent with what Martinez (2012) found for US school data, where the addition of the classroom level reduced the variance of both the pupil and the school level.

If the classroom level is decidedly kept in the models for accountability, this begs the question of who should be accountable for these effects. A simple route to take is to hold the teachers accountable; however, this is rather simplistic and ultimately unfair, insofar as teachers could be made responsible for the effects of streaming within their school. As with the case of the school type variable discussed in Chapter 3, the choice of retaining the classroom level in a model for school accountability is highly debatable. However, the stance of this thesis is that schools should not necessarily be rewarded for effective streaming practices, even at the expense of potentially obscuring part of the effects truly proceeding from better school practices, and hence the classroom level should be kept and explored more deeply, via school inspection or qualitative inquiry, for instance.

On a side note, as discussed in Chapter 2, CVA models can be used for mainly two reasons: school accountability and school choice. About the latter, Leckie and Goldstein (2009) have argued that estimated CVA scores have important limitations to inform parents' school choice. According to them, standard CVA estimates do not take the uncertainty of predicting future school performance into account and hence they are useless for informing parents. They show that when this uncertainty is accounted for by computing standard errors in a non-standard way, the confidence intervals of the estimated future CVA scores are so wide that hardly any school can be distinguished from the national average or even from any other school. Even though the main purpose of this thesis is not related to school choice, the limitation described here with regard to future school performance does affect the models in this thesis. Nevertheless, as Leckie and Goldstein (2011b) point out, performance tables (or accountability measures in this thesis) derived from these models are still retrospectively informative and provide highly relevant information for holding schools accountable.

On another front, some authors (Foley & Goldstein, 2012; Leckie & Goldstein, 2011b) have argued that the meaning and limitations of CVA scores (school effects estimates used in England) and their confidence intervals have not been properly disseminated to the public, which prevents them from properly interpreting and using this information. In order to minimise the downsides of using CVA scores, Leckie and Goldstein (2011b) propose a new methodology to make multiple tailor-made school comparisons to inform parents' choice. This methodology involves simulating the probabilities of pairs of (or several) schools to be ranked higher than the other(s), thus making school performance information more accessible to a non-statistical audience, because of the familiarity with the concept of probability in the wider public. This methodology has not been implemented in this thesis; however, it is certainly an avenue for further research that is worthwhile exploring.

With respect to the above mentioned, the school classifications derived from the CVA models fitted in Chapters 4 and 5 can be regarded as deterministic, insofar as they only admit one classification per school, while in reality this can be much more uncertain. Nevertheless, the use of confidence intervals for the accountability measures (school residuals from the extended CVA models) makes this method more reliable than simply using raw averages as current practice in Chile is (still) set up to do.

In sum, results from this thesis show that the way in which a CVA model is specified is a sensitive matter, which can have significant policy implications for diverse external stakeholders and serious consequences for the schools and the pupils. As discussed in Chapter 2, the new Chilean Education Quality Assurance System foresees the utilisation of the school classifications for supporting schools with insufficient performance; however, persistent insufficient performance may result in closure (Agencia de Calidad de la Educación, 2014; San Martín & Carrasco, 2013). The choice of methods is therefore crucial.

7.4. The importance of internal school accountability

In this thesis, the concept of cultural capital (Bourdieu, 1977) was chosen as a source of insight into the matter of informing “internal” or “intelligent” school accountability, as described by San Martín and Carrasco (2013) and Sahlberg (2007, 2010). Theoretically, cultural capital is thought to be rooted in educational inequalities due to the socio-economic gradient. Given the profound socio-economic disparities in Chile, as analysed in Chapters 4 and 5, the use of the concept of cultural capital renders even more relevant.

Authors such as Edgerton et al. (2012), Edgerton and Roberts (2014), and Sullivan (2001, 2002) have argued in favour of the use of Bourdieuvian concepts as means of disentangling persistent educational inequalities. Nevertheless, the diversity of evidence regarding the effect of cultural capital on academic performance depends heavily on the operationalisation of the concept. As seen in Chapter 6 of this thesis, cultural capital is latent variable estimated from a measurement model that includes manifest variables related to reading habits, access to books and parental qualifications as other authors (Sullivan 2001, 2002) have recommended for their usefulness.

The latent construct "cultural capital" has been found to have a sizable and non-linear effect on performance in Language, whereas its effect on Mathematics, controlling for the same set of variables, is also considerable. Although there are some differences in how cultural capital affects performance in Spanish Language and Mathematics, especially at the extreme ends of the distribution, the overall effect follows the same pattern in both subjects: "culturally-advantaged" pupils tend to outperform the "culturally-disadvantaged", which is consistent with a Bourdieuvian theoretical perspective. Other potentially relevant routes for further analyses can be to investigate possible differences across time between various cohorts of pupils and differences across regions or localities in the effect of cultural capital on academic performance. In line with Sullivan (2002), the effect of cultural capital can be hypothesised to vary through time and across countries; even though this has not been explored in this thesis, the potential differences across different contexts have empirical support as it was found that the effect of cultural capital varies by schools.

Finally, from a substantive point of view, this thesis shows how access to all sorts of reading materials and reading habits can have not only a relevant impact on pupils' progress in Language, but also in Mathematics. This thesis advocates for re-thinking school accountability as other authors have proposed (San Martín & Carrasco, 2013; Sahlberg 2007, 2010) and shows that CVA models can be used to inform internal school accountability and potentially serve as drivers of school improvement policies, provided the models reasonably control for known non-malleable factors (as seen in Chapters 4 and 5) to satisfy the demands of external school accountability and additionally analyse the effect of meaningful malleable factors, such as cultural capital as seen in Chapter 6. Ultimately, this can potentially constitute useful information for implementing policy not only at the school level, but also at the level of the local and central government offices.

In the next chapter, the conclusions of this thesis are presented. Consistent with the ideas discussed in this chapter, the conclusions are discussed around three main ideas: firstly, CVA models for school accountability, either external or internal, need to take into account the complexity of influences affecting pupils' academic progress as thoroughly as possible in order to make a fair assessment of schools' performance and/or to inform school improvement policies accurately. Secondly, school effectiveness is not a unidimensional process, which implies that CVA models should ideally (when data are available) reflect upon the multidimensionality of academic performance and take into consideration the relationship between different subjects as well as non-academic outcomes. Thirdly, CVA models can also be used to inform internal school accountability by analysing the effects of meaningful modifiable factors and potentially serve as drivers of school improvement policies.

Chapter 8: Conclusions

8.1. Statement of the conclusions

This research had set out to analyse school value-added and pupils' academic progress in Chile with the ultimate purpose of providing insight into the development of models of external and internal school accountability. The main conclusions arising from this research are summarised below:

Firstly, contextualised value-added models for school accountability, either external or internal, need to take into account the complex network of influences affecting pupils' academic progress as thoroughly as possible in order to make a fair assessment of schools' performance and/or to inform school improvement policies. More specifically, this requires the implementation of models accounting for sources of variation in academic progress beyond the schools and the pupils, i.e. the effects of classrooms within schools, the carry-over effects of primary schools, the geographical effects of local authorities and others, should data availability allow it.

Secondly, school effectiveness is not a unidimensional process, which implies that contextualised school value-added models should (depending on data availability) reflect upon the multidimensionality of the phenomenon. This would convey taking into consideration the relationship between different subjects and non-academic outcomes. This should be done ideally at all levels, but at least at the pupil and the school level. Such models are undoubtedly richer and have the potential of informing school accountability more reliably, insofar as they allow a more comprehensive judgment of the overall contribution of the school to a more unabridged educational development of its pupils.

Thirdly, contextualised value-added models can also be used to inform internal school accountability and potentially serve as drivers of school improvement policies, provided the models reasonably control for known non-malleable factors and analyse the effect of meaningful malleable factors. The value of such models lies in that they have the potential to contribute to the understanding of how to overcome or at least circumvent, via school or public policy, the seemingly overpowering effect of the socio-economic background on academic performance.

These conclusions, especially the first two, follow the basic principle that school effectiveness models should be sufficiently and realistically complex to hold schools accountable externally as fairly as possible. This is also one of the guiding principles for the third conclusion, although from the perspective of internal school accountability, where the main concern is to distinguish between school or pupil factors on which the school can realistically intervene to improve its own practices and achieve higher goals.

8.2. Significance and contribution to knowledge

The purpose of this section is to revisit and briefly respond to the research questions, making a recapitulation of the main empirical findings to discuss their contribution to knowledge.

- What are the effects of classrooms, local authorities and primary schools on Mathematics and Spanish Language progress in Chile, beyond pupils' socio-economic and demographic characteristics and secondary school-level effects?

Academic progress is a complex phenomenon influenced by many external and internal factors. As discussed throughout this thesis, school value-added models have customarily assessed pupils' academic progress controlling for pupils' heterogeneity and between-school variation. Even though this approach has been proved successful in identifying educational inequalities and helping to assess schools more fairly, its main shortcoming is that it does not control for relevant factors external to the pupils and the schools, which renders such models incomplete for external and internal school accountability.

In Chapters 4 and 5, this research has implemented multilevel models with increased complexity compared to the traditional school value-added approach, to demonstrate that the differences between classrooms within secondary schools, between primary schools and between local authorities constitute significant sources of variation in the progress that pupils in Chilean schools make in Mathematics and Spanish Language. Furthermore, it has been demonstrated that a considerable proportion of the total variation in pupils' attainment and progress does not actually proceed from the pupils themselves nor the secondary school they attend. This occurs even after controlling for the carry-over effects of primary school, socio-economic and demographic characteristics at both the school level and the pupil level.

- How are pupils' progress and school value-added in Mathematics related to pupils' progress and school value-added in Spanish Language in Chile?

Another caveat of the traditional approach to study academic performance is that it does not take the multidimensionality of the phenomenon into account. As discussed extensively earlier in this thesis, implementing univariate school value-added models implies assuming that pupils split up their learning process into different unrelated areas of knowledge and that schools break down the teaching process into different dissociated subjects. This is certainly an untenable assumption for three main reasons. While it is true that the curriculum is divided into several components, making connections between subjects is not prevented but encouraged. Furthermore, even in the case in which the connections between subjects were neglected in a school or in the hypothetical case in which they were actively discouraged, the very plausible existence of shared teaching practices would still have common effects across subjects at the aggregated school level. Additionally, pupils are taught and learn knowingly or unknowingly a number of transferrable skills throughout their educational trajectories, which makes the relationships between subjects even more plausible.

In Chapter 5, this research used a bivariate multilevel model to demonstrate that there is an underlying positive relationship between progress in Mathematics and progress in Spanish Language at the levels of pupils, classrooms, primary schools, secondary schools and local authorities. As expected, this relationship varies in magnitude, but not in direction, across levels. Focusing on the pupils, the underlying relationship between Mathematics and Spanish Language is considerable when fitting an unconditional 2-level model, but only moderate after controlling for all relevant levels of variation and for socio-economic and demographic characteristics of the pupils themselves and the schools. This implies that a great deal of this relationship is due to confounding. However, after confounders are removed, the remainder would seem to indicate that while pupils can be more inclined or able in one subject or another, they can still learn throughout their educational trajectories a set of skills (although undetermined in this research) that are applicable to both subjects. With regard to the schools, the underlying relationship between both subjects is quite sizable, which implies that secondary schools making significant contributions to their pupils' progress in one subject are also very likely to be doing the same in the other subject (and others as well). The opposite, of course, would also be true. Focusing on what school policy can intervene, this could be said to indicate that teaching practices shared within secondary schools are somewhat uniformly effective or ineffective irrespective of the subject.

- Are there any relevant (non-negligible) differences between school accountability measures derived from diverse statistical models?

School value-added models have been used extensively to derive measures to inform school accountability externally to various stakeholders and to inform parents' choice of school. As exposed and discussed thoroughly in this thesis, these models suffer from serious shortcomings that are mainly the consequence of misspecification. More specifically, traditional school value-added models are misspecified inasmuch as they do not control for relevant non-malleable factors; hence, accountability measures derived from them make schools responsible, for better or worse, for circumstances over which they have no control.

In Chapters 4 and 5 of this thesis, it has been demonstrated that the choice of statistical model is not trivial, because large differences were found between school accountability measures derived from a traditional 2-level contextualised value-added model and the extended 4 and 5-level contextualised value-added models. Controlling for the additional levels of classrooms, primary schools (only included in Chapter 5) and local authorities revealed that the effects of schools and pupils' heterogeneity had been largely overestimated when fitting models from a traditional approach. The huge overestimation of the contribution of schools to pupils' progress in Mathematics and Spanish Language is apparent in the accountability measures derived from the extended value-added models, which distinguished from the national average fewer than half of the schools that the traditional approach distinguished. Considering that the extended value-added models fitted in this research account for more non-malleable factors external to the control of the schools and the pupils (or their parents and carers) than the traditional models, one would be inclined to judge the school accountability measures from the extended CVA model as more accurate and ultimately fairer.

- How does cultural capital along with socio-economic and demographic characteristics of the pupils and the schools affect pupils' progress and school effectiveness in Chile?

School value-added models are not solely circumscribed to fulfil the purpose of external accountability. As discussed earlier in this thesis, school effectiveness models can be used to provide insight into academic performance to a variety of stakeholders, either external, such as policy-makers, local and central government officials, tax-payers, parents choosing schools, etc., or internal, such as head teachers, school governors, teachers and the pupils themselves, provided that their particular needs of information are met. This requires tailoring the statistical models to ensure practical adequacy. In general, external accountability models require comprehensive control for non-malleable factors affecting pupils' progress to ensure fair comparisons. In contrast, internal school accountability models imply focusing on malleable school and pupil factors to inform school improvement interventions, as well as controlling for the non-malleable factors to avoid confounding. With these principles in mind, cultural capital was targeted in this research as a malleable factor potentially susceptible of school intervention which can ultimately contribute to better educational outcomes.

In Chapter 6 of this thesis, it was found that cultural capital, a pupil-level latent variable materialised in reading habits, number of books possessed and parental qualifications, had a sizable effect on pupils' progress in Spanish Language and Mathematics. Given that the measurement model for such latent variable was focused mainly on reading-related variables, the greater impact was observed, as expected, on Language; however, its effect is nonetheless relevant on Mathematics progress as well. Given that cultural capital can be thought of as being intrinsically limited, non-linear effects were fitted and proved significant; however, substantive differences were observed for the case of progress in Language as compared to Mathematics. While pupils with extreme values of cultural capital have a shallower rate of progress in Mathematics, the curvilinear effect of cultural capital on Language shows that small changes in cultural capital at the extremes can prompt more pronounced boosts in pupils' progress. In practical terms, this would reinforce the idea that interventions involving the improvement of reading habits or book accessibility, for instance, may potentially enhance pupils' progress.

With respect to the socio-economic and demographic characteristics of the schools and the pupils, their effects as non-malleable factors were analysed extensively throughout Chapters 4, 5 and 6. By including these variables in the models, it was possible to ascertain a number of educational inequalities with respect to gender, income, school socio-economic status and school institutional type. The purpose (and value) of controlling for these variables is twofold: on the one hand, they allow distinguishing between what can be modified or not by the schools or the pupils themselves and, on the other hand, they allow tailoring school interventions to suit the needs of particular groups of pupils.

The overall contribution to knowledge of this thesis is related to three main aspects. Firstly, from an international perspective, this thesis confirms once more that the school performance phenomenon is indeed complex, as shown by other studies in developed countries, such as the United Kingdom, the Netherlands and the United States. Furthermore, it shows that the challenges that this complexity conveys are even more marked in developing countries, such as Chile. As seen in this thesis, the size of the school effects and the geographical effects are far greater in Chile than in developed countries, which makes the use of sophisticated methods even more relevant. In sum, assessing school performance in developing countries needs to be tackled with advanced modelling approaches to fully capture its complexity.

Secondly, from a methodological point of view, the methods described and applied in this thesis are certainly not new and have been applied in other studies; however, none of the studies reviewed had applied all the different specifications presented here at once. For instance, as mentioned in chapter 2, a few researchers have pointed out the need for specifying further levels to the traditional 2-level structure of pupils nested within schools; some of those studies have incorporated the analysis of fully-nested structures, and some others have included the analysis of non-nested structures. On another front, a very small number of studies have specified multivariate multilevel models for analysing educational outcomes. Nevertheless, no other studies have specified further nested and non-nested levels (beyond the traditional 2-level structure) in a multivariate multilevel model to analyse educational outcomes. This thesis has demonstrated that this avenue, even though complex, is fruitful for analysing school performance and deriving fairer school accountability measures for policy-making purposes.

Thirdly, in the Chilean public policy context, this thesis contributes to the debate about how to assess school performance more fairly. As seen throughout this thesis, the models presented are certainly an improvement with respect to the traditional approach of 2-level CVA models of pupils nested within schools, let alone with respect to the new policy of the Chilean Government. This thesis clearly shows that the choice of models is not trivial, especially when considering that an unfair/incomplete assessment can potentially have unduly harsh consequences on the schools and the pupils themselves.

8.3. Theoretical implications

Empirical analyses throughout this thesis have shown the value of incorporating several variables into the models of school value-added and pupils' academic progress. Naturally, the incorporation of such variables follows the examples and general principles of the theoretical knowledge built upon previous research. For instance, theories have established the widely-known effects of certain variables such as socio-economic status of the schools and the pupils, demographic variables such as gender and ethnicity, etc. In this regard, this research included several of these variables as controls to ensure fair school comparisons, along with other external factors. A number of special considerations regarding models of school effectiveness with theoretical repercussions can be derived from the conclusions of this thesis.

By concluding that academic progress is a complex multidimensional phenomenon influenced by many agents and contexts at multiple levels, it is possible to identify several components to be taken into account.

With respect to the complexity of academic progress, models should mimic this complexity, as far as possible and within reason, to ensure a fair and realistic representation of the underlying processes resulting in pupils' progress.

Regarding their multidimensionality, models should reflect upon the fact that pupils' progress, as well as school processes and practices, emerge in multiple interrelated academic subjects and non-academic outcomes.

Moreover, from a multilevel perspective, the model building process should not only include the effects proceeding from the pupils and the schools, but also further effects proceeding from classrooms, primary schools, local authorities and other relevant external factors affecting pupils' progress. This last consideration is to be pondered regardless of the stakeholders to be informed.

From a multi-agency perspective, models should always be constructed bearing in mind the audience or stakeholders to be informed as the main guiding principle for the inclusion of malleable or non-malleable variables in the analyses. On the one hand, if the models are constructed for informing school accountability to external stakeholders, then the inclusion of key non-malleable school and pupil factors ought to be carefully thought, taking into consideration the theory and the context in which the school system operates. On the other hand, if the models are constructed to inform internal school accountability, then theory should determine the inclusion of substantively meaningful malleable school and pupil factors susceptible of intervention.

A more specific considerations in the building process of models for external school accountability is the inclusion of certain key school-level non-malleable factors. Models should control, as far as possible, for school selectivity in an attempt to counter the bias arising from the ill-conceived school policy of improving results via manipulation of the pupil intake. Models should also include institutional type as a non-malleable factor affecting pupils' progress, even at the expense of partially obscuring effects truly proceeding from diverse school practices associated with school type, because not all schools have the same degree of freedom to modify their own institutional regime. This is especially relevant in countries with highly segregated school systems such as Chile.

With respect to malleable factors affecting pupils' progress, this research demonstrated the substantive and practical value of the theoretical concept of cultural capital, when building a model for internal school accountability. From a theoretical perspective, the significance of this finding is threefold: firstly, it showed that culturally-disadvantaged pupils are less likely to make as much progress as their culturally-advantaged counterparts; secondly, it showed that cultural capital has a contextual component, which is reflected on the fact that pupils can benefit in terms of academic progress from an overall school-level cultural capital higher than their own; and thirdly, it showed that cultural capital has the potential of affecting progress in multiple subjects and not only language.

The implications and considerations discussed above are not meant to constitute an exhaustive list, nor are they exclusively related to theoretical issues. In practice, these issues also reach the field of public policy and they should be pondered in policy-making processes, as detailed next.

8.4. Policy implications

The implications of this research in terms of public policy have already been outlined in the previous section and discussed throughout previous chapters. These are mainly related to the design, implementation and use of school accountability models and measures either with the purpose of informing external stakeholders or internal stakeholders.

Government practice in Chile prior to 2015 consisted basically on using school averages as indicators of effectiveness, with the purpose of informing external school accountability and parental choice indiscriminately. The shortcomings of such an approach have been thoroughly discussed in Chapter 2 of this thesis. Policy-making processes guided by said approach run into several problems: firstly, they convey the risk of ignoring differential effects on certain groups of pupils, who might potentially be in disadvantage. Secondly, school averages are potentially misleading, because they overestimate the overall contribution of schools to their pupils' academic progress by ignoring relevant external influential factors. Thirdly, school averages unadjusted for pupils' prior attainment do not allow for the comparison of schools based on the progress that their pupils make. They can only assess the level of attainment at certain points in time (cross-sectionally), which may eventually convey a severe misjudgement of the contribution of schools. This could potentially result in an incentive for schools to manipulate their pupil intake instead of actually improving their practices to achieve better results.

As mentioned before in Chapter 2, a new school accountability system has been set out by the Chilean Ministry of Education via the Agency for the Quality of Education and it is due to be implemented in a pilot stage in 2015. This quality assurance system will hold schools accountable based on their pupils' scores obtained in the SIMCE tests. Schools will be classified in four categories of performance (insufficient, lower-middle, middle and high) adjusting for a few socio-economic variables and prior attainment (where available). Even though, some lessons learned from school value-added research have been considered in this system, a rather inconvenient and invalid method has been chosen to be applied to make these adjustments, i.e. Multiple Linear Regression. The main caveats of such approach were discussed thoroughly in Chapters 2 and 3.

This research demonstrated that ignoring certain influential factors can have severe consequences in the assessment of schools. An accountability system, either to inform external or internal stakeholders, whose main driving purpose is fairness to all schools and pupils, cannot disregard the effects of the classrooms over which neither the parents or the pupils have control to choose; the effects of primary schools over which secondary schools have no control to intervene; or the effects of the local authorities where pupils reside and schools are inserted. These effects could certainly be controlled for in a Multiple Linear

Regression approach, as proposed by the Chilean Agency for the Quality of Education, but such a task would be unnecessarily cumbersome and it might still run the risk of drawing mistaken conclusions. This in turn would lead to imprecise school classifications, which can eventually have stern repercussions for the schools themselves.

A direct and precise comparison between the models presented in this thesis and the models developed by the Chilean Agency for the Quality of Education is not possible, because the proposed accountability system is still in a pilot stage and details of the methodology have not been made available to the public yet. Nevertheless, considering that the models presented in chapters 4 and 5 of this thesis have demonstrated that the traditional 2-level CVA models considerably overestimate the school effects, one can only assume that the school accountability measures derived from MLR models will overestimate these effects even further. Put simply, this thesis advocates for a better use of the available information on school performance to improve the process of evaluation of schools. From what it is known and foreseeable given the released information, the new Government policy on school accountability can potentially be unfairly detrimental to many schools. In short, this policy needs to be revised and improved.

In sum, the main practical lesson learned from this thesis is that models for external school accountability need to account for the available information about the classrooms, primary schools and local authorities, as well as the relationship between diverse subjects taught in schools, to ensure fairer comparisons. This has also proved to be effective even for models whose aim is to inform internal school improvement processes. From the perspective of internal school accountability, this research has demonstrated that policy changes towards increasing accessibility to reading materials (fiction and non-fiction books, magazines on general culture or specialised topics, etc.) and improving reading habits can have a significant impact on progress in Language and even in Mathematics. Although this is not proved in this thesis, the fact that the effect of cultural capital is significant not only on Language, but also on Mathematics makes it very plausible that it has a significant effect on subjects such as Science, History, etc.

This by no means implies that the models presented in this thesis are infallible; however, they constitute a significant contribution towards a more comprehensive analysis of the school performance phenomenon in Chile.

8.5. Limitations of the study

As hinted previously, this research is not free from limitations. The first and foremost limitation of this research is related to data availability, which is reflected on two main aspects.

Firstly, despite having pointed out the multidimensionality of academic progress, this research only implemented models for progress in Mathematics and Spanish Language, simply because no other longitudinal test data were available at the time of this research. However, the latest reforms to the educational quality assurance system foresee the application of tests in other subjects longitudinally, as well as the collection of data regarding non-academic outcomes.

The second main caveat of this research is that the stability over time of the estimated school effects has not been studied. This is again related to data availability, because, as mentioned before in Chapter 3 of this thesis, the cohort under study (2004-2006) is the only one moving from the last year of primary school to secondary school that was available at the time of this research. Nonetheless, as is the case of the first main caveat, the latest reforms can remedy this situation and in a few years' time the SIMCE information system will be much more enriched.

On another front, as discussed in chapter 3, it goes without saying the importance of adequately handling the missingness in statistical modelling. However, the complexity of the models and the size of the data sets can certainly be an obstacle. In this research, one of the most important measures to be analysed are the school residuals, which were used to derive school accountability measures. For convenience, complete case analyses were carried out in all stages of this research, since the computational burden and complexity outweighed the potential benefits of conducting multiple imputation. A somewhat similar approach was adopted in previous studies (see for example: Leckie, 2009; Rasbash et al., 2010). Moreover, Gelman et al. (2005) recognise the impracticality of post-estimation procedures with multiply imputed data sets and, especially, in latent factor analysis and multilevel settings. Nevertheless, the authors describe and discuss methods to carry out post-estimation procedures, making full use of multiply imputed data sets. This could potentially be a fruitful avenue to explore in further work.

Despite these limitations, the analyses presented in this thesis have been carried out rigorously and have demonstrated practical adequacy to support its value. Moreover, the conclusions arising from these analyses can serve as a guide for future research, tackling the present limitations with further data to come in subsequent years and build upon their strengths.

8.6. Recommendations for future research

It has been underlined previously that the extensions to the traditional school value-added models presented in this research are not negligible and not trivial whatsoever. Hence, further research on school value-added in Chile needs to incorporate at least information on classrooms, primary schools and local authorities, as well as outcomes from diverse subjects. Failing to do so would resonate on the reliability of the estimated school effects.

Prospectively, further models could be built up from the models presented in this thesis introducing additional data to come in subsequent years. Firstly, information from multiple cohorts could help improve the reliability and precision of the estimation of school effects by analysing their stability or change over time. This would be especially useful for an accountability system, because it would contribute to ascertain which schools are moving across classifications, either upwards or downwards. An additional source of insight might be gained through implementing a probabilistic approach to school classifications as discussed in Chapter 7. Eventually, these processes of change could also be further investigated via qualitative research.

Secondly, information from additional multiple outcomes could be used to ascertain further relationships between subjects and eventually non-academic outcomes. This would certainly contribute to provide a wider picture of pupils' progress and school value-added, which can be purposefully used for planning and implementing school interventions. At the school or classroom level, it would most likely allow a closer examination of the results of shared teaching practices and collaboration across subjects. At the pupil-level, such analyses could eventually contribute to a deeper knowledge of the learning process, which in turn can be used as valuable insight for curriculum planning. This insight could potentially be enriched with further information about teachers; however, this conveys the obstacle of the SIMCE databases not allowing the linkage between teachers and specific classrooms within secondary schools. This is an issue that should be resolved in subsequent waves of the SIMCE tests. Furthermore, an investigation of the relationship between non-academic outcomes (which are not necessarily related to school practices) and achievement of the curriculum goals, which are more likely to be related to school practices, might also be of special interest for qualitative inquiry.

Thirdly, regarding internal school accountability, further research could benefit with additional data on malleable pupil-level and school-level factors. In this research, the effect of the latent variable cultural capital was analysed; however, the particular specification of the measurement model could eventually be modified or extended to include more variables related to reading and study habits, as well as parental involvement, should such data become available.

On a side note, future research would certainly benefit from tackling issues surrounding missing data. As mentioned in the previous section, Gelman et al. (2005) describe and discuss methods to carry out post-estimation procedures to make full use of multiply imputed data sets.

Lastly, there are other possible avenues for further research regarding connections beyond those between multiple subjects. Given the importance of the context where the learning process takes place, connections between pupils as social networks are potentially a fruitful choice for closer investigation. This is supported by the evidence found in this research that cultural capital also has a relevant contextual component. Moreover, connections between schools can also be thought of as relevant factors influencing school effectiveness; the underlying mechanism could be via shared human and financial resources between schools within private consortia or public associations. The combination of all these potential factors would require the adoption of a multilevel social network analysis approach.

8.7. Final remarks

As can be seen from the previous discussions, the analysis of academic performance is multifaceted and may seem inextricable at times. Nevertheless, the challenge that this task poses can be taken up gradually if approached with appropriate research questions, theories and methods. This research demonstrated that the current sophistication of the statistical tools and its foreseeable future sophistication, along with the ever-increasing availability of data and computational capacity, allow pushing the boundaries of standard approaches and tackling the complexity to achieve a better understanding of the phenomenon. This can ultimately yield benefits not only for the pupils and the schools, but also for Society as a whole.

Ultimately, the overall contribution and significance of this research lies in the fact that it demonstrates the methodological and substantive value of extending the traditional approach of school value-added. For the methodological point of view, this research showed how the extensions are valid and useful for analysing progress in Mathematics and Language, either for the purpose of external school accountability or school improvement. From the substantive point of view, this thesis showed that these extensions are useful tools for better informed public policy. Finally, this research contributes to a deeper knowledge of how the school effectiveness phenomenon in the context of an unequal developing country such as Chile, differs with respect to how it takes shape in more egalitarian developed countries.

References

Agencia de Calidad de la Educación. (2014). *Minuta informe metodología de ordenación [Notes on report of ranking methodology]*. Santiago, Chile. Retrieved from http://www.agenciaeducacion.cl/

Anand, S., & Sen, A. (1994). *Human development index: Methodology and measurement*. New York, US. Retrieved from http://hdr.undp.org/sites/default/files/oc12.pdf

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, *36*(5), 258–267.

Bernstein, B. (1975). *Class, Codes and Control: Towards A Theory Of Educational Transmissions* (Vol. 3). London, UK: Routledge & Kegan Paul.

Berrington, A., Smith, P., & Sturgis, P. (2006). *An Overview of Methods for the Analysis of Panel ESRC National Centre for Research Methods NCRM Methods Review Papers*. Southampton. Retrieved from http://eprints.ncrm.ac.uk/415/1/MethodsReviewPaperNCRM-007.pdf

Bourdieu, P. (1977a). Cultural reproduction and social reproduction. In J. Karabel & A. Halsey (Eds.), *Power and ideology in education*. New York, US: Oxford University Press.

Bourdieu, P. (1977b). *Outline of a theory of practice*. (J. Goody, Ed.)*Cambridge studies in social anthropology* (Vol. 16). Cambridge, UK: Cambridge University Press. Retrieved from http://www.loc.gov/catdir/description/cam022/76011073.html

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of Theory of Research for the Sociology of Education* (pp. 241–258). New York, US: Greenwood Press.

Bourdieu, P. (1988). *Homo academicus*. Cambridge, UK: Polity Press.

Bourdieu, P. (1990). *The logic of practice*. Cambridge, UK: Polity Press.

Bourdieu, P., & Passeron, J. (1977). *Reproduction in education, society and culture*. London and Beverly Hills: SAGE Publications.

Bourdieu, P., & Passeron, J. (1979). *The inheritors: French students and their relation to culture*. Chicago, US: University of Chicago Press.

Bourdieu, P., & Wacquant, L. (1992). *An Invitation to Reflexive Sociology*. Chicago, US: University of Chicago Press.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. London, UK: Routledge and Kegan Paul.

Bowles, S., & Gintis, H. (2002). Schooling in capitalist America. *Sociology of Education*, *75*(1), 1–18.

Browne, W. (2012). MCMC Estimation in MLwiN, v2.26. Centre for Multilevel Modelling, University of Bristol.

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

Byrne, D., & Rogers, T. (2000). Divided Spaces - Divided School: An Exploration of the Spatial Relations of Social Division. *Sociological Research Online*, *5*(3), XV–XVI.

Carpenter, J., Goldstein, H., & Kenward, M. (2011). REALCOM-IMPUTE Software for multilevel multiple imputation with mixed response types. *Journal of Statistical Software*, *45*(5), 1–12.

Castells, M. (1996). *The rise of the network society*. Oxford, UK: Blackwell Publishers Ltd.

Centro de Estudios MINEDUC. (2014). *Estadísticas de la Educación 2013*. Santiago, Chile. Retrieved from http://centroestudios.mineduc.cl/index.php?t=96&i=2&cc=2036&tm=2

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

Cervini, R. (2009b). Comparando la inequidad en los logros escolares de la educación primaria y secundaria en Argentina: Un estudio multinivel [Comparing inequity of educational achievement in primary and secondary education in Argentina: A multilevel study]. *Revista Iberoamericana Sobre Calidad, Eficacia Y Cambio En Educación (REICE)*, *7*(1), 5–21.

Coe, R., & Fitz-Gibbon, C. (1998). School effectiveness research: criticisms and recommendations. *Oxford Review of Education*, *24*(4), 421–438.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC, US Government Printing Office.

Cowan, R., Donlan, C., Newton, E., & Lloyd, D. (2005). Number Skills and Knowledge in Children With Specific Language Impairment. *Journal of Educational Psychology*, *97*(4), 732–744.

Creemers, B. (1994). The history, value and purpose of school effectiveness studies. In D. Reynolds, B. Creemers, P. Nesselrodt, E. Schaffer, S. Stringfield, & C. Teddlie (Eds.), *Advances in school effectiveness research and practice* (pp. 9–23). Oxford, UK: Elsevier Science.

Creemers, B. (2007). Educational effectiveness and improvement: the development of the field in mainland Europe. In T. TOWNSEND (Ed.), *International handbook of school effectiveness and improvement* (Vol. 1). The Netherlands: Springer.

Cuthbert, R. (2010). Students as Customers? *Higher Education Review*, *42*(3), 3–25.

de Ayala, R. (2013). *Theory and Practice of Item Response Theory*. New York: Guilford Publications. Retrieved from https://books.google.com/books?id=NchXAQAAQBAJ&pgis=1

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

De Graaf, P. (1986). The Impact of Financial and Cultural Resources on Educational Attainment in the Netherlands. *Sociology of Education*, *59*, 237–246.

De Leeuw, J., & Meijer, E. (2008). Introduction to multilevel analysis. In J. De Leuuw & E. Meijer

(Eds.), *Handbook of multilevel analysis* (pp. 1–75). New York: Springer.

Deming, E. (1993). *The New Economics For Industry, Government and Education*. Cambridge, Massachusetts, US: Massachusetts Institute of Technology Press.

Di Maggio, P. (1982). Cultural Capital and School Success. *American Sociological Review*, *47*, 189–201.

Di Maggio, P., & Mohr, J. (1985). Cultural Capital, Educational Attainment, and Marital Selection. *American Journal of Sociology*, *90*, 1231–1261.

Dobson, A. (2002). *An introduction to generalized linear models, 2nd edition* (2nd ed.). Boca Raton, London, New York, Washington, D.C.: Chapman & Hall/CRC.

Donlan, C., Cowan, R., Newton, E., & Lloyd, D. (2007). The role of language in mathematical development: evidence from children with specific language impairments. *Cognition*, *103*(1), 23–33.

Dumais, S. (2002). Cultural Capital, Gender, and School Success: The Role of Habitus. *Sociology of Education*, *75*, 44–68.

Duncan, O., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, *20*, 200–217.

Durkheim, E. (1947). *The division of labor in society / translated from the French by George Simpson*. (G. Simpson, Ed.). Glencoe, Illinois, US: Free Press.

Durkheim, E. (1956). *Education and sociology / translated and with an introduction by Sherwood D. Fox ; foreword by Talcott Parsons*. New York, US and London, UK: Free Press and Collier Macmillan.

Ebbes, P., Böckenholt, U., & Wedel, M. (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica*, *58*(2), 161–178.

Edgerton, J., & Roberts, L. (2014). Cultural capital or habitus? Bourdieu and beyond in the explanation of enduring educational inequality. *Theory and Research in Education*, *12*(2), 193–220. doi:10.1177/1477878514530231

Edgerton, J., Roberts, L., & Peter, T. (2012). Disparities in Academic Achievement: Assessing the Role of Habitus and Practice. *Social Indicators Research*, *114*(2), 303–322.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(1), 15–27.

Emery, C., Kramer, T., & Tian, R. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, *11*(1), 37–46.

European Foundation for Quality Management. (2015). The EFQM excellence model. Retrieved January 25, 2015, from http://www.efqm.org/the-efqm-excellence-model

Feinstein, L., Duckworth, K., & Sabates, R. (2004). *A model of the intergenerational transmission of educational success*. London: The Centre for Research on the Wider Benefits of Learning, Institute of Education. Retrieved from http://www.learningbenefits.net/Publications/ResRepIntros/ResRep10intro.htm

Ferrão, M., & Goldstein, H. (2009). Adjusting for measurement error in the value added model: Evidence from Portugal. *Quality & Quantity*, *43*(6), 951–963.

Fielding, A., & Goldstein, H. (2006). *Cross-classified and Multiple Membership Structures in Multilevel Models : An Introduction and Review*. (DFES, Ed.). London: Department for Education and Skills. Retrieved from https://www.education.gov.uk/publications/standard/publicationDetail/Page1/RR791

Fielding, A., Thomas, H., Steele, F., Browne, W., Leyland, A., Spencer, N., & Davison, I. (2006). *Using Cross-classified multilevel models to improve estimates of the determination of pupil attainment: A scoping study*. Birmingham.

Finney, S., & DiStefano, C. (2006). Non-normal and categorical data in Structural Equation Modeling. In G. Hancock & R. Mueller (Eds.), *Structural Equation Modeling: a second course* (pp. 269–314). Greenwich, US: Information Age Publishing.

Fitz-Gibbon, C. (1997). *The value added national project: Final report*. London: SCAA.

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

Franz, R. (1998). Whatever you do, don't treat your students like customers! *Journal of Management Education*, *22*(1), 63–69.

Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D., & Meulders, M. (2005). Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics*, *61*, 74–85.

Gibson, A., & Asthana, S. (1998). Schools, Pupils and Examination Results: contextualising school "performance." *British Educational Research Journal*, *24*(3), 269–282.

Giddens, A. (1993). *Sociology*. Cambridge, UK: Polity Press.

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

Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: Scope and limitations. *British Educational Research Journal*, *27*(4), 433–442.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: John Wiley and Sons, Ltd.

Goldstein, H., Browne, W., & Rasbash, J. (2002). *Partitioning variation in multilevel models*. *Understanding Statistics*. London, UK: Institute of Education. Retrieved from http://seis.bris.ac.uk/~frwjb/materials/pvmm.pdf

Goldstein, H., Burgess, S., & McConnell, B. (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.

Goldstein, H., & Healy, M. (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., & Leckie, G. (2008). School league tables: what can they really tell us?

Significance, *5*(2), 67–69.

Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *159*(3), 385–443.

Goldstein, H., & Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *159*(1), 149–163.

Halpern, D. (2005). *Social Capital*. Cambridge, UK: Polity Press.

Hancock, G., & Mueller, R. (Eds.). (2006). *Structural Equation Modeling: a second course*. Greenwich, US: Information Age Publishing.

Hargreaves, D. (1994). *Changing teachers, changing times: teachers' work and culture in the post-modern age*. London, UK: Cassell.

Hecht, S., Torgesen, J., Wagner, R., & Rashotte, C. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: a longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*(2), 192–227.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Hernandez-Martinez, P., Black, L., Williams, J., Davis, P., Pampaka, M., & Wake, G. (2008). Mathematics students' aspirations for higher education: class, ethnicity, gender and interpretative repertoire styles. *Research Papers in Education*, *23*(2), 153–165.

Hides, M., Davies, J., & Jackson, S. (2004). Implementation of EFQM excellence model self-assessment in the UK higher education sector - lessons learned from other sectors. *The TQM Magazine*, *16*(3), 194–201.

Hill, P., & Goldstein, H. (1998). Multilevel Modeling of Educational Data with Cross-Classification and Missing Identification for Units. *Journal of Educational and Behavioural Statistics* , *23*(2), 117–128.

Hill, P., & Rowe, K. (1996). Multilevel Modelling in School Effectiveness Research. *School Effectiveness and School Improvement*, *7*(1), 1–34.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hox, J. (2010). *Multilevel analysis: Techniques and applications, 2nd edition*. *Quantitative methodology series* (2nd ed.). New York and Hove, East Sussex: Routledge.

Illich, I. (1973). *Deschooling society*. Harmondsworth, UK: Penguin.

International Organization for Standardization. (2015). ISO 9000 - Quality management. Retrieved January 25, 2015, from http://www.iso.org/iso/iso_9000

Isaki, C., & Fuller, W. (1982). Survey Design Under the Regression Superpopulation Model. *Journal of the American Statistical Association*, *77*(377), 89–96.

Izadi, M., Kashef, A., & Stadt, R. (1996). Quality in Higher Education: Lessons Learned from the Baldrige Award, Deming Prize, and ISO 9000 Registration. *Journal of Industrial Teacher Education*, *33*(2).

Jencks, C., Smith, M., Acland, H., Bane, M., Cohen, D., Gintis, H., … Michelson, S. (1972). *Inequality: a reassessment of the effect of family and schooling in America*. New York, US: Basic Books.

Kaplan, D. (2009). *Structural Equation Modeling (2nd ed.): Foundations and Extensions.* Thousand Oaks, USA: SAGE Publications.

Kingston, P. (2001). The Unfulfilled Promise of Cultural Capital Theory. *Sociology of Education*, (Extra issue), 88–89.

Kleanthous, I., & Williams, J. (2013). Perceived parental influence and students' dispositions to study mathematically-demanding courses in Higher Education. *Research in Mathematics Education*, *15*(1), 50–69. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/14794802.2013.763608

Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, London: The Guilford Press.

Kumar, K., & Sarangapani, P. (2004). History of the Quality Debate. *Contemporary Education Dialogue*, *2*, 30–52.

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.

Lareau, A., & Weininger, E. (2003). Cultural capital in educational research: A critical assessment. *Theory and Society*, *32*, 567–606.

Lareau, A., & Weininger, E. (2004). Cultural capital in educational research. In D. Swartz & V. Zolberg (Eds.), *After Bourdieu: Influence, critique, elaboration* (pp. 105–144). Dordrecht, The Netherlands: Kluwer academic publishers.

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., & Charlton, C. (2013). runmlwin : A Program to Run the MLwiN Multilevel Modeling Software from within Stata. *Journal of Statistical Software*, *52*(11).

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(4), 835–851.

Leckie, G., & Goldstein, H. (2011a). A note on "The limitations of school league tables to inform school choice." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(3), 833–836.

Leckie, G., & Goldstein, H. (2011b). Understanding uncertainty in school league tables. *Fiscal Studies*, *32*(2), 207–224.

Leckie, G., Pillinger, R., Jenkins, J., & Rasbash, J. (2010). School, family, neighbourhood: which is most important to a child's education? *Significance*, *7*(2), 67–70.

Lingard, B., Ladwig, J., & Luke, A. (1998). School effects in postmodern conditions. In R. Slee, G. Weiner, & S. Tomlinson (Eds.), *School effectiveness for whom? Challenges to the school effectiveness and school improvement movements* (pp. 84–100). London, UK: Falmer Press.

Lundquist, R. (1997). Quality Systems and ISO 9000 in Higher Education. *Assessment & Evaluation in Higher Education*, *22*(2), 159–172.

Luyten, H., Visscher, A., & Witziers, B. (2005). School Effectiveness Research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, *16*(3), 249–279.

Manzi, J., San Martín, E., & Van Bellegem, S. (2014). School system evaluation by value added analysis under endogeneity. *Psychometrika*, *79*(1), 130–153.

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.

McEwan, P. (2001). The effectiveness of public, Catholic, and non-religious private schools in Chile's voucher system. *Education Economics*, *9*(2), 25.

McEwan, P., & Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis*, *22*(3), 26.

Michael, R., Sower, V., & Motwani, J. (1997). A comprehensive model for implementing total quality management in higher education. *Benchmarking for Quality Management & Technology*, *4*(2), 104–120.

Ministry of Education - Chile. (2015). Nivel de rendimiento: 2006 al 2010. Retrieved from http://datos.gob.cl/datasets/ver/990

Mizala, A., & Romaguera, P. (2000). School performance and choice: The Chilean experience. *The Journal of Human Resources*, *35*(2), 392–417.

Mizala, A., & Romaguera, P. (2001). Factores Socioeconómicos explicativos de los resultados escolares en la educación secundaria en Chile. *El Trimestre Económico2*, *68*(272), 515–549.

Mizala, A., & Torche, F. (2012). Bringing the schools back in: the stratification of educational achievement in the Chilean voucher system. *International Journal of Educational Development*, *32*(1), 132–144.

Munoz-Chereau, B. (2013). *Searching for fairer ways of comparing Chilean secondary schools performance: a mixed methods study investigating contextual value added approaches.* University of Bristol.

Murillo, F. J., & Román, M. (2011). School infrastructure and resources do matter: analysis of the incidence of school resources on the performance of Latin American students. *School Effectiveness and School Improvement*, *22*(1), 29–50.

doi:10.1080/09243453.2010.543538

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*, 551–560.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 43–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

Muthén, L., & Muthén, B. (2012). *Mplus User's Guide. 7th edition*. Los Angeles, US: Muthén and Muthén.

OECD. (2008). *Measuring Improvements in Learning Outcomes. Best practices to assess the value-added of schools*. Paris: OECD.

ONS. (2006). Review of the Dissemination of Health Statistics: Confidentiality Guidance. (O. for N. Statistics, Ed.). London.

Paredes, R., & Pinto, J. (2009). ¿El fin de la educación pública en Chile? Is this the end of public education in Chile? *Estudios de Economía*, *36*(1), 47–66.

Plewis, I. (1996). Statistical methods for understanding cognitive growth : A review , a synthesis and an application. *British Journal of Mathematical and Statistical Psychologycal Psychology*, *49*, 25–42.

Plewis, I. (2011). Contextual variations in ethnic group differences in educational attainments. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(2), 419–437. Retrieved from http://doi.wiley.com/10.1111/j.1467-985X.2010.00679.x

Rasbash, J., Charlton, C., Browne, W., Healy, M., & Cameron, B. (2012). MLwiN version 2.26 [Computer software]. Bristol, UK: Centre for Multilevel Modelling, University of Bristol.

Rasbash, J., Leckie, G., & Pillinger, R. (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.

Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2012). A User's Guide to MLwiN, v2.26. Centre for Multilevel Modelling, University of Bristol.

Raudenbush, S. (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.

Ray, A. (2006). *School value added measures in England*. London. Retrieved from https://education.gov.uk/publications/eOrderingDownload/RW85.pdf

Rea, J., & Weiner, G. (1998). Cultures of blame and redemption-When empowerment becomes control: practitioners' views of the effective schools movement. In R. Slee, G. Weiner, & S. Tomlinson (Eds.), *School effectiveness for whom? Challenges to the school effectiveness and school improvement movements* (pp. 21–32). London, UK: Falmer Press.

Reynolds, D. (1995). The effective school: an inaugural lecture. *Evaluation and Research in Education*, *9*(2), 57–73.

Reynolds, D., & Teddlie, C. (2001). Reflections on the critics and beyond them. *School Effectiveness and School Improvement*, *12*, 99–113.

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., & Townsend, T. (2000). An introduction to school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research*. New York, US and London, UK: Falmer Press.

Robbins, D. (2005). The origins, early development and status of Bourdieu's concept of "cultural capital." *The British Journal of Sociology*, *56*(1), 13–30.

Robinson, W. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, *15*(3), 351–357.

Sahlberg, P. (2007). Education policies for raising student learning: the Finnish approach. *Journal of Education Policy*, *22*(2), 147–171.

Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, *11*, 45–61.

Sallis, E. (2002). *Total Quality Management in Education* (3rd ed.). London: Kogan Page Ltd.

Sammons, P., Nuttall, D., & 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., & Mortimore, P. (1997). *Forging links. Effective schools and effective departments*. London: Paul Chapman Publishing Ltd.

San Martín, E., & 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 mo. *Temas de La Agenda Pública*, *7*(53), 1–18.

San Martín, E., & Carrasco, A. (2013). Criterios para evaluar la metodología oficial de clasificación de escuelas: ¿un asunto técnico o conceptual? [Criteria to assess the official methodology of classification of schools: a technical or conceptual matter?]. In I. Irarrázaval, M. Morandé, & M. Letelier (Eds.), *Propuestas para Chile [Proposals for Chile]* (pp. 85–114). Centro de Políticas Públicas, Pontificia Universidad Católica de Chile. Retrieved from http://politicaspublicas.uc.cl/wp-content/uploads/2014/01/Libro-Propuestas-para-Chile_versión-web.pdf

Sax, B. (2004). Students as "customers." *On the Horizon*, *12*(4), 158–160.

Scheerens, J. (2000). *Improving school effectiveness*. *Fundamentals of educational planning*. Paris: UNESCO - International Institute for Educational Planning. Retrieved from http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED459535

Scheerens, J. (2004). Review of school and instructional effectiveness research. Paper commissioned for the Education For All (EFA) Global Monitoring Report 2005, The Quality Imperative. Retrieved from http://unesdoc.unesco.org/images/0014/001466/146695e.pdf

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. London, UK:

Pergamon.

Simmons, F. R., & Singleton, C. (2008). Representations Impact on Review of Research into Arithmetic and Dyslexia. *Dyslexia*, *14*, 77–94.

Simonite, V., & Browne, W. (2003). Estimation of a large cross–classified multilevel model to study academic achievement in a modular degree course. *Ournal of the Royal Statistical Society, Series A*, *166*(1), 119–134.

Snijders, T., & Bosker, R. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modelling, 2nd edition*. London, Thousand Oaks, New Delhi: SAGE Publications.

Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

StataCorp. (2011). Stata Statistical Software: Release 12. College Station, Texas.

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *171*(1), 5–19.

Stoll, L., & Mortimore, P. (1997). School effectiveness and school improvement. In M. Barber & J. White (Eds.), *Perspectives on school effectiveness and school improvement*. London: Institute of Education.

Sulis, I., & Porcu, M. (2014). Assessing Divergences in Mathematics and Reading Achievement in Italian Primary Schools: A Proposal of Adjusted Indicators of School Effectiveness. *Social Indicators Research*, 1–28.

Sullivan, A. (2001). Cultural Capital and Educational Attainment. *Sociology*, *35*(4), 893–912.

Sullivan, A. (2002). Bourdieu and education: How useful is Bourdieu's theory for researchers? *The Netherlands' Journal of Social Sciences*, *38*(2), 144–166.

Tate, W. (1997). Race-Ethnicity , SES , Gender , and Language Proficiency Trends in Mathematics Achievement: An Update. *Journal for Research in Mathematics Education*, *28*(6), 652–679.

The World Bank. (2011). Poverty reduction and equity: Measuring inequality. Retrieved from http://go.worldbank.org/3SLYUTVY00

The World Bank. (2013). World Development Indicators: GINI Index. Retrieved from http://data.worldbank.org/indicator/SI.POV.GINI

Timmermans, A. (2012). *Value added in educational accountability: Possible, fair and useful?* University of Groningen, Groningen, The Netherlands.

Timmermans, A., Doolaard, S., & 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., Snijders, T., & Bosker, R. (2013). In search of value added in the case of complex school effects. *Educational and Psychological Measurement*, *73*(2), 210–228.

Tokman, A. (2002). *Is private education better? Evidence from Chile*. *Central Bank of Chile Working Papers* (Vol. 147). Santiago, Chile. Retrieved from http://www.bcentral.cl/eng/studies/working-papers/147.htm

Torche, F. (2005). Privatization reform and inequality of educational opportunity: The case of Chile. *Sociology of Education*, *78*, 316–343.

Tranmer, M., & Steel, D. (2001). Ignoring a level in a multilevel model: evidence from UK census data. *Environment and Planning*, *33*, 941–948.

UNESCO. (2004). *Education for all. The quality imperative.* Paris: UNESCO Publishing. Retrieved from http://unesdoc.unesco.org/images/0013/001373/137333e.pdf

UNICEF. (2000). *Defining quality in education*. *UNICEF Working Paper Series*. New York, US. Retrieved from http://www.unicef.org/education/files/QualityEducation.PDF

United Nations Development Programme. (2004). *Las trayectorias del Desarrollo Humano en las comunas de Chile [The trajectories of human development in the communes of Chile]*. Santiago, Chile. Retrieved from http://www.desarrollohumano.cl/otraspub/pub12/IDHC con portada.pdf

United Nations Development Programme. (2013). *Human development report 2013. The rise of the South: Human progress in a diverse world. Explanatory note on 2013 HDR composite indices*. New York, US. Retrieved from http://hdr.undp.org/sites/default/files/Country-Profiles/CHL.pdf

Valenzuela, J., Bellei, C., & De los Ríos, D. (2008). *Evolución de la Segregación Socioeconómica de los Estudiantes Chilenos y su Relación con el Financiamiento Compartido*. *Proyecto FONIDE - Fondo de Investigación y Desarrollo en Educación*. Santiago, Chile. Retrieved from http://www.facso.uchile.cl/psicologia/epe/_documentos/GT_cultura_escolar_politica_educativa/recursos bibliograficos/articulos relacionados/valenzuelabelledelosros(2009)segregacionyficom.pdf

Van Den Berghe, W. (1998). Application of ISO 9000 Standards to Education and Training. *Vocational Training: European Journal*, *15*, 20–28.

Van Landeghem, G., De Fraine, B., & Van Damme, J. (2010). The consequence of ignoring a level of nesting in Multilevel Analysis: A comment. *Multivariate Behavioral Research*, *40*(4), 424–434.

Vukovic, R., & Lesaux, N. (2013). The relationship between linguistic skills and arithmetic knowledge. *Learning and Individual Differences*, *23*, 87–91.

Wang, J., & Goldschmidt, P. (1999). Opportunity to Learn , Language Proficiency , and Immigrant Status Effects on Mathematics Achievement. *The Journal of Educational Research*, *93*(2), 101–111.

Williams, J. (2012). Use and exchange value in mathematics education: Contemporary CHAT meets Bourdieu's sociology. *Educational Studies in Mathematics*, *80*, 57–72.

Williams, J., Black, L., Hernandez-Martinez, P., Davis, P., Pampaka, M., & Wake, G. (2009).

Repertoires of aspiration, narratives of identity, and cultural models of mathematics in practice. In M. Cesar & K. Kumpulainen (Eds.), *Social interactions in multicultural settings* (pp. 39–69). Rotterdam, The Netherlands: Sense Publishers.

Willis, P. (1977). *Learning to labour: how working class kids get working class jobs*. London, UK: Saxon House.

Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. University of California. Retrieved from http://www.statmodel.com/download/Yudissertation.pdf

Appendices

Appendix 1: Complementary and intermediate models from Chapter 4

1.1 Fixed-effects parameters of the CVA model for progress in Mathematics without controlling for cross-level interaction effects.

Main effects†	Coef.	S. E.	z	P>z	95% Conf. Int.	
Intercept	-0.309	0.015	-20.515	0.000	-0.338	-0.279
Prior attainment	0.598	0.004	160.252	0.000	0.590	0.605
Male	0.069	0.005	15.200	0.000	0.060	0.078
Lower-middle income	0.018	0.004	4.129	0.000	0.010	0.027
Upper-middle income	0.031	0.007	4.331	0.000	0.017	0.044
High income	0.035	0.010	3.454	0.001	0.015	0.056
Held back	-0.259	0.008	-33.613	0.000	-0.275	-0.244
Lower-middle SES	0.108	0.018	6.113	0.000	0.073	0.142
Middle SES	0.305	0.020	15.138	0.000	0.266	0.345
Upper-middle SES	0.532	0.024	21.912	0.000	0.484	0.579
Upper SES	0.724	0.046	15.584	0.000	0.633	0.816
Subsidised independent	0.055	0.015	3.716	0.000	0.026	0.085
Independent	0.027	0.043	0.617	0.537	-0.058	0.111
Interaction effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Prior att. & Male	-0.010	0.003	-3.290	0.001	-0.017	-0.004
Prior att. & Low-mid income	0.009	0.003	2.695	0.007	0.002	0.016
Prior att. & Up-mid income	0.015	0.005	2.984	0.003	0.005	0.025
Prior att. & High income	0.020	0.007	3.043	0.002	0.007	0.033
Prior att. & Held back	-0.071	0.005	-13.800	0.000	-0.081	-0.061
Male & Low-mid income	-0.012	0.006	-1.887	0.059	-0.024	0.000
Male & Up-mid income	-0.017	0.009	-1.794	0.073	-0.036	0.002
Male & High income	0.002	0.012	0.138	0.891	-0.021	0.025
Male & Held back	0.081	0.009	8.915	0.000	0.063	0.098

Note: The random part and the model fit information are displayed in Tables 4.10 and 4.11, respectively.
† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.

1.2. Fixed-effects parameters of the CVA model for progress in Mathematics controlling for cross-level interaction effects and including non-significant interactions between prior attainment and income.

Main effects†	Coef.	S. E.	z	P>z	95% Conf. Int.	
Intercept	-0.318	0.015	-20.573	0.000	-0.348	-0.288
Prior attainment	0.582	0.006	95.356	0.000	0.570	0.594
Male	0.069	0.005	15.064	0.000	0.060	0.077
Lower-middle income	0.018	0.004	4.045	0.000	0.009	0.027
Upper-middle income	0.032	0.007	4.503	0.000	0.018	0.046
High income	0.036	0.011	3.411	0.001	0.015	0.056
Held back	-0.258	0.008	-33.398	0.000	-0.273	-0.243
Lower-middle SES	0.113	0.018	6.194	0.000	0.077	0.149
Middle SES	0.330	0.021	15.900	0.000	0.289	0.371
Upper-middle SES	0.549	0.025	22.228	0.000	0.500	0.597
Upper SES	0.749	0.047	15.991	0.000	0.657	0.840
Subsidised independent	0.063	0.015	4.133	0.000	0.033	0.092
Independent	0.028	0.043	0.645	0.519	-0.057	0.113
Interaction effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Prior att. & Male	-0.011	0.003	-3.339	0.001	-0.017	-0.004
Prior att. & Low-mid income‡	0.002	0.004	0.557	0.577	-0.005	0.009
Prior att. & Up-mid income‡	0.003	0.005	0.625	0.532	-0.007	0.014
Prior att. & High income‡	0.009	0.008	1.163	0.245	-0.006	0.025
Prior att. & Held back	-0.067	0.005	-12.946	0.000	-0.077	-0.057
Male & Low-mid income‡	-0.012	0.006	-1.930	0.054	-0.024	0.0002
Male & Up-mid income‡	-0.017	0.009	-1.823	0.068	-0.036	0.001
Male & High income‡	0.001	0.012	0.117	0.907	-0.022	0.024
Male & Held back	0.081	0.009	8.898	0.000	0.063	0.098
Prior attainment & Low-mid SES	0.0002	0.007	0.024	0.981	-0.013	0.013
Prior attainment & Middle SES	0.043	0.008	5.451	0.000	0.028	0.059
Prior attainment & Up-mid SES	0.046	0.010	4.667	0.000	0.027	0.066
Prior attainment & Upper SES	0.011	0.021	0.546	0.585	-0.029	0.052
Prior attainment & Subs. independent	0.013	0.005	2.506	0.012	0.003	0.024
Prior attainment & Independent	0.020	0.019	1.051	0.293	-0.017	0.057

† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.

‡ These interactions became non-significant after specifying the cross-level interaction effects, and hence they were removed from subsequent models.

1.3. Fixed-effects parameters of the CVA model for progress in Mathematics controlling for cross-level interaction effects and including non-significant interactions between gender and school characteristics.

Main effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Intercept	-0.317	0.016	-19.813	0.000	-0.348	-0.286
Prior attainment	0.581	0.006	96.833	0.000	0.569	0.593
Male	0.067	0.008	8.375	0.000	0.051	0.083
Lower-middle income	0.017	0.005	3.4	0.001	0.007	0.027
Upper-middle income	0.030	0.007	4.286	0.000	0.016	0.044
High income	0.034	0.011	3.091	0.002	0.012	0.056
Held back	-0.258	0.008	-32.25	0.000	-0.274	-0.242
Lower-middle SES	0.109	0.019	5.737	0.000	0.072	0.146
Middle SES	0.331	0.021	15.762	0.000	0.290	0.372
Upper-middle SES	0.550	0.025	22	0.000	0.501	0.599
Upper SES	0.746	0.049	15.224	0.000	0.650	0.842
Subsidised independent	0.064	0.015	4.267	0.000	0.035	0.093
Independent	0.032	0.045	0.711	0.477	-0.056	0.120
Interaction effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Prior attainment & Male	-0.010	0.003	-3.333	0.001	-0.016	-0.004
Prior attainment & Low-mid income	0.002	0.004	0.5	0.617	-0.006	0.010
Prior attainment & Up-mid income	0.003	0.005	0.6	0.549	-0.007	0.013
Prior attainment & High income	0.009	0.008	1.125	0.261	-0.007	0.025
Prior attainment & Held back	-0.067	0.005	-13.4	0.000	-0.077	-0.057
Male & Low-mid income‡	-0.010	0.006	-1.667	0.096	-0.022	0.002
Male & Up-mid income‡	-0.013	0.010	-1.3	0.194	-0.033	0.007
Male & High income‡	0.006	0.015	0.4	0.689	-0.023	0.035
Male & Held back	0.080	0.009	8.889	0.000	0.062	0.098
Male & Low-mid SES‡	0.012	0.010	1.200	0.230	-0.008	0.032
Male & Middle SES‡	-0.004	0.012	-0.333	0.739	-0.028	0.020
Male & Up-mid SES‡	-0.004	0.016	-0.25	0.803	-0.035	0.027
Male & Upper SES‡	0.008	0.035	0.229	0.819	-0.061	0.077
Male & Subs. Indep‡.	-0.006	0.008	-0.75	0.453	-0.022	0.010
Male & Independent‡	-0.011	0.032	-0.344	0.731	-0.074	0.052
Prior attainment & Low-mid SES	-0.001	0.007	-0.143	0.886	-0.015	0.013
Prior attainment & Middle SES	0.043	0.008	5.375	0.000	0.027	0.059
Prior attainment & Up-mid SES	0.046	0.010	4.6	0.000	0.026	0.066
Prior attainment & Upper SES	0.011	0.021	0.524	0.600	-0.030	0.052
Prior attainment & Subs. independent	0.014	0.005	2.8	0.005	0.004	0.024
Prior attainment & Independent	0.020	0.019	1.053	0.293	-0.017	0.057

† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.
‡ These interactions became non-significant after specifying the cross-level interaction effects, and hence they were removed from subsequent models.

1.4. Fixed-effects parameters of the traditional 2-level CVA model for progress in Mathematics used to compare with the extended 4-level CVA model of chapter 4.

Main effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Intercept	-0.274	0.015	-18.405	0.000	-0.303	-0.245
Prior attainment	0.622	0.006	99.345	0.000	0.610	0.634
Male	0.055	0.004	14.685	0.000	0.048	0.062
Lower-middle income	0.017	0.003	5.081	0.000	0.010	0.024
Upper-middle income	0.031	0.005	5.772	0.000	0.020	0.041
High income	0.049	0.008	6.190	0.000	0.033	0.064
Held back	-0.289	0.008	-36.374	0.000	-0.305	-0.274
Lower-middle SES	0.081	0.018	4.519	0.000	0.046	0.117
Middle SES	0.263	0.020	13.214	0.000	0.224	0.302
Upper-middle SES	0.470	0.024	19.997	0.000	0.424	0.516
Upper SES	0.663	0.045	14.787	0.000	0.575	0.751
Subsidised independent	0.034	0.015	2.203	0.028	0.004	0.064
Independent	0.009	0.042	0.223	0.824	-0.073	0.092
Interaction effects†	**Coef.**	**S. E.**	**z**	**P>z**	**95% Conf. Int.**	
Prior att. & Male	-0.010	0.003	-3.099	0.002	-0.016	-0.004
Prior att. & Held back	-0.082	0.005	-15.500	0.000	-0.093	-0.072
Male & Held back	0.084	0.009	9.032	0.000	0.066	0.102
Prior att. & Low-mid SES	-0.002	0.007	-0.343	0.732	-0.017	0.012
Prior att. & Middle SES	0.025	0.008	3.098	0.002	0.009	0.041
Prior att. & Up-mid SES	0.020	0.010	2.048	0.041	0.001	0.039
Prior att. & Upper SES	-0.018	0.021	-0.874	0.382	-0.059	0.023
Prior att. & Subs. independent	-0.018	0.006	-3.022	0.003	-0.030	-0.006
Prior att. & Independent	-0.006	0.020	-0.303	0.762	-0.045	0.033

† Reference categories: Female; Low income; Not held back; State-funded school and Low school SES.

1.5. Random-effects parameters of the traditional 2-level CVA model for progress in Mathematics used to compare with the extended 4-level CVA model of chapter 4.

Parameter					
Level 2: School	**Estimate**	**S.E.**	**Correlation**	**95% Coverage int.**	
Variance (Intercept)	0.074	0.003	--	0.069	0.079
Covariance (Prior attainment, Intercept)	0.006	0.001	0.284	0.004	0.007
Variance (Prior attainment)	0.006	0.0004	--	0.005	0.007
Covariance (Male, Intercept)	0.001	0.001	0.064	-0.001	0.003
Covariance (Male, Prior attainment)	-0.001	0.0004	-0.158	-0.002	-0.0001
Variance (Male)	0.005	0.001	--	0.004	0.007
Level 1: Pupil	**Estimate**	**S.E.**	**Correlation**	**95% Coverage int.**	
Variance (Intercept)	0.306	0.001	--	0.304	0.308
Model fit information					
Deviance	308863.1				
AIC	308921.1				
Number of parameters	29				

Appendix 2: Complementary and intermediate models from Chapter 5

2.1. Fixed-effects parameters from the CVA bivariate model for progress in Mathematics and Language, including cross-level interaction effects and the non-significant interaction between prior attainment and income.

Fixed effects Mathematics†					Fixed effects Language†				
Pupil-level main effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level main effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Intercept	-0.343	0.016	-0.373	-0.312	Intercept	-0.234	0.012	-0.258	-0.210
Prior attainment	0.551	0.006	0.540	0.562	Prior attainment	0.615	0.006	0.605	0.626
Male	0.079	0.005	0.070	0.088	Male	-0.032	0.005	-0.042	-0.023
Lower-middle income	0.017	0.004	0.008	0.025	Lower-middle income	0.033	0.005	0.023	0.042
Upper-middle income	0.032	0.007	0.018	0.046	Upper-middle income	0.063	0.008	0.047	0.079
High income	0.040	0.010	0.019	0.060	High income	0.061	0.012	0.038	0.085
Held back	-0.269	0.008	-0.284	-0.254	Held back	-0.219	0.008	-0.236	-0.203
School-level main effects§	**Coef.**	**S.E.**	**95% C.I.**		**School-level main effects§**	**Coef.**	**S.E.**	**95% C.I.**	
Subs. Indep. school	0.059	0.016	0.028	0.090	Subs. Indep. school	0.049	0.012	0.026	0.073
Independent school	0.023	0.045	-0.066	0.111	Independent school	0.024	0.036	-0.048	0.095
Low-mid school SES	0.104	0.019	0.067	0.141	Low-mid school SES	0.071	0.014	0.043	0.099
Middle school SES	0.320	0.022	0.278	0.363	Middle school SES	0.255	0.017	0.222	0.287
Up-mid school SES	0.554	0.026	0.503	0.605	Up-mid school SES	0.437	0.020	0.397	0.477
Upper school SES	0.777	0.049	0.683	0.873	Upper school SES	0.624	0.040	0.546	0.701
Pupil-level interaction effects‡	**Coef.**	**S.E.**	**95% C.I.**		**Pupil-level interaction effects‡**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Male	-0.006	0.003	-0.012	0.000	Prior att. & Male	-0.049	0.003	-0.056	-0.043
Prior att. & Low-mid inc.	0.002	0.003	-0.005	0.009	Prior att. & Low-mid inc.	-0.007	0.004	-0.015	0.000
Prior att. & Up-mid inc.	0.004	0.005	-0.006	0.014	Prior att. & Up-mid inc.	0.004	0.006	-0.008	0.016
Prior att. & High inc.	0.010	0.008	-0.006	0.025	Prior att. & High inc.	0.003	0.009	-0.015	0.021
Prior att. & Held back	-0.068	0.005	-0.078	-0.058	Prior att. & Held back	-0.069	0.005	-0.080	-0.059
Male & Low-mid inc.	-0.013	0.006	-0.025	-0.001	Male & Low-mid inc.	-0.004	0.007	-0.017	0.010
Male & Up-mid inc.	-0.020	0.009	-0.038	-0.001	Male & Up-mid inc.	-0.013	0.011	-0.034	0.008
Male & High inc.	-0.003	0.012	-0.026	0.020	Male & High inc.	0.019	0.013	-0.006	0.044
Male & Held back	0.084	0.009	0.066	0.102	Male & Held back	0.029	0.010	0.009	0.049
Cross-level interaction effects‡§	**Coef.**	**S.E.**	**95% C.I.**		**Cross-level interaction effects‡§**	**Coef.**	**S.E.**	**95% C.I.**	
Prior att. & Subs. Indep.	0.012	0.005	0.002	0.022	Prior att. & Subs. Indep.	-0.001	0.005	-0.010	0.009
Prior att. & Indep school	0.014	0.018	-0.022	0.049	Prior att. & Indep school	0.031	0.019	-0.006	0.069
Prior att. & Low-mid SES	-0.011	0.006	-0.023	0.002	Prior att. & Low-mid SES	-0.022	0.006	-0.034	-0.010
Prior att. & Middle SES	0.021	0.007	0.006	0.035	Prior att. & Middle SES	-0.012	0.007	-0.026	0.002
Prior att. & Up-Mid SES	0.020	0.009	0.001	0.038	Prior att. & Up-Mid SES	-0.009	0.009	-0.027	0.009
Prior att. & Upper SES	-0.022	0.020	-0.060	0.017	Prior att. & Upper SES	-0.054	0.021	-0.094	-0.013

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 300,000; burn-in: 500; storing all iterations. All fixed-effects parameters have an effective sample size of at least 5,500.

‡ Reference categories for pupil-level variables: Female; Low income and Not held back.

§ Reference categories for school-level variables: State-funded school and Low school SES.

2.2. Extended MCMC information of the fixed-effects parameters from full bivariate 5-level CVA model for progress in Mathematics and Spanish Language.

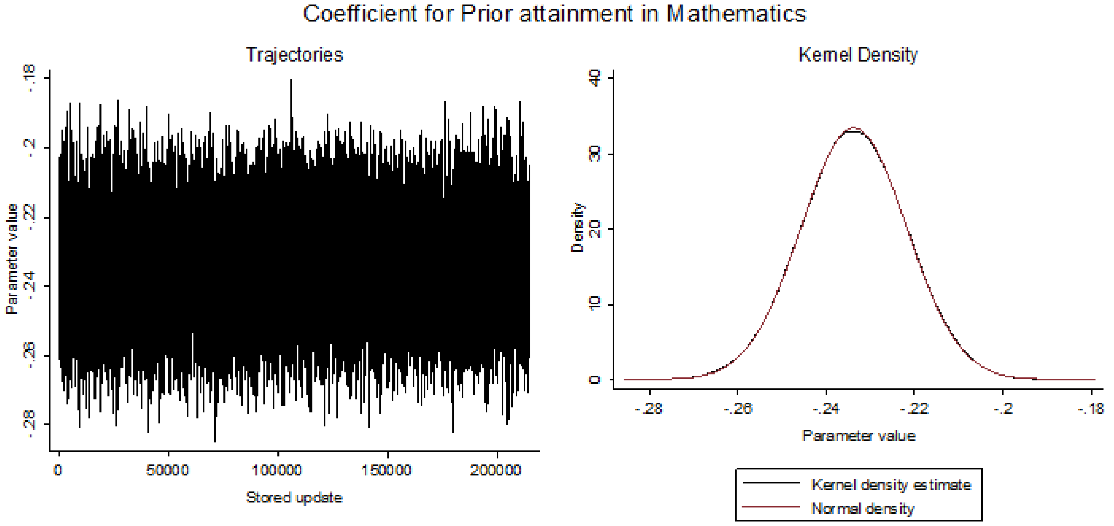
Fixed effects Mathematics†						
Pupil-level main effects‡	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Intercept	-0.342	0.016	-0.342	-0.372	-0.312	5733
Prior attainment	0.551	0.006	0.551	0.540	0.563	61229
Male	0.079	0.005	0.079	0.070	0.088	87751
Lower-middle income	0.016	0.004	0.016	0.008	0.025	174113
Upper-middle income	0.032	0.007	0.032	0.018	0.046	171654
High income	0.044	0.010	0.044	0.025	0.062	154029
Held back	-0.269	0.008	-0.269	-0.284	-0.254	182095
School-level main effects§	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Subsidised Indep. school	0.059	0.016	0.059	0.029	0.090	3815
Independent school	0.022	0.045	0.022	-0.067	0.110	9825
Lower-middle school SES	0.104	0.019	0.104	0.067	0.141	4447
Middle school SES	0.320	0.022	0.320	0.278	0.363	4120
Upper-middle school SES	0.554	0.026	0.554	0.503	0.605	4907
Upper school SES	0.775	0.049	0.775	0.680	0.872	9738
Pupil-level interaction effects‡	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Prior att. & Male	-0.006	0.003	-0.006	-0.012	0.0002	114684
Male & Low-mid income	-0.013	0.006	-0.013	-0.025	-0.001	170486
Male & Up-mid income	-0.019	0.009	-0.019	-0.038	-0.001	167423
Male & High income	-0.002	0.012	-0.002	-0.024	0.021	132734
Prior att. & Held back	-0.068	0.005	-0.068	-0.078	-0.058	177124
Male & Held back	0.084	0.009	0.084	0.066	0.102	186244
Cross-level interaction effects‡§	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Prior att. & Subs. Indep.	0.012	0.005	0.012	0.002	0.022	35358
Prior att. & Indep. school	0.015	0.018	0.015	-0.020	0.050	71151
Prior att. & Low-mid SES	-0.010	0.006	-0.010	-0.023	0.002	55481
Prior att. & Middle SES	0.022	0.007	0.022	0.008	0.036	52990
Prior att. & Up-Mid SES	0.023	0.009	0.023	0.005	0.040	55033
Prior att. & Upper SES	-0.016	0.019	-0.016	-0.053	0.021	73601
Fixed effects Language†						
Pupil-level main effects‡	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Intercept	-0.234	0.012	-0.234	-0.257	-0.210	7959
Prior attainment	0.615	0.006	0.615	0.604	0.626	81059
Male	-0.033	0.005	-0.032	-0.042	-0.023	104299
Lower-middle income	0.032	0.005	0.032	0.022	0.041	182518
Upper-middle income	0.067	0.008	0.067	0.051	0.082	183570
High income	0.065	0.011	0.065	0.044	0.086	167981
Held back	-0.220	0.008	-0.220	-0.236	-0.203	188638
School-level main effects§	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Subsidised Indep. school	0.049	0.012	0.049	0.026	0.072	5030
Independent school	0.023	0.036	0.023	-0.048	0.094	14208
Lower-middle school SES	0.071	0.015	0.071	0.043	0.100	6088
Middle school SES	0.254	0.017	0.254	0.222	0.287	5509
Upper-middle school SES	0.436	0.020	0.436	0.396	0.475	6573
Upper school SES	0.622	0.039	0.622	0.544	0.700	13710
Pupil-level interaction effects‡	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Prior att. & Male	-0.050	0.003	-0.050	-0.056	-0.043	133007
Male & Low-mid income	-0.002	0.007	-0.002	-0.016	0.012	179897
Male & Up-mid income	-0.014	0.011	-0.014	-0.035	0.007	176131
Male & High income	0.019	0.013	0.019	-0.006	0.044	146857
Prior att. & Held back	-0.069	0.005	-0.069	-0.080	-0.058	182983
Male & Held back	0.029	0.010	0.029	0.009	0.049	188431
Cross-level interaction effects‡§	**Coef.**	**S.E.**	**Median**	**95% C.I.**		**ESS**
Prior att. & Subs. Indep.	-0.001	0.005	-0.001	-0.010	0.009	51727
Prior att. & Indep. school	0.033	0.019	0.033	-0.004	0.070	105111
Prior att. & Low-mid SES	-0.023	0.006	-0.023	-0.035	-0.011	74535
Prior att. & Middle SES	-0.014	0.007	-0.014	-0.028	-0.001	71617
Prior att. & Up-Mid SES	-0.010	0.009	-0.010	-0.026	0.007	76865
Prior att. & Upper SES	-0.052	0.020	-0.052	-0.091	-0.013	105912

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 215,000; burn-in: 500; storing all iterations. Trajectories are depicted for the shaded parameter values only as they are representative of all the estimated fixed-effects parameters. More details are given below. ESS=Effective sample size.

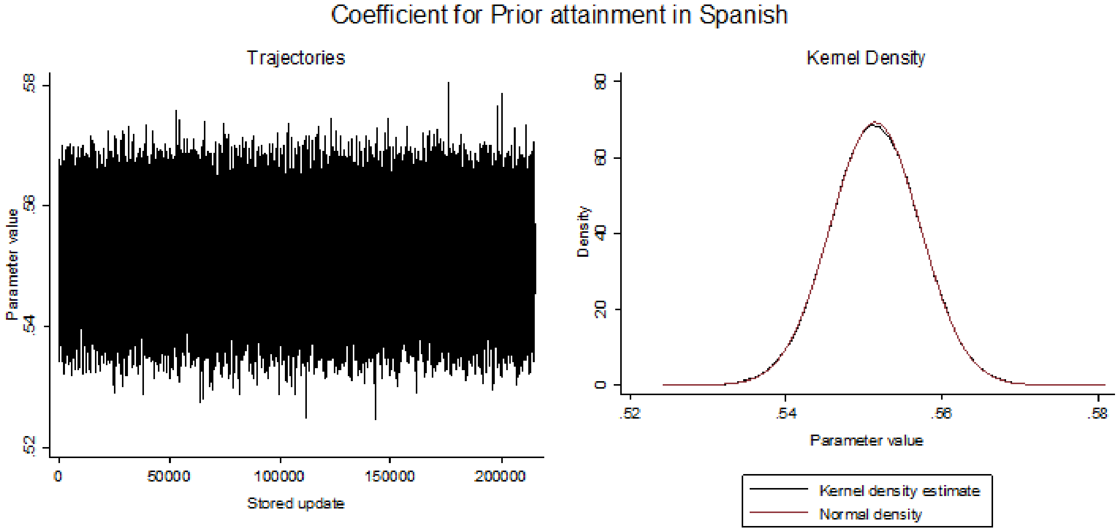
‡ Reference categories for pupil-level variables: Female; Low income and Not held back.

§ Reference categories for school-level variables: State-funded school and Low school SES.

2.3. Trajectories and Density of the fixed-effect parameter for prior attainment in Mathematics from the full bivariate CVA model for progress in Mathematics and Spanish Language.



2.4. Trajectories and Density of the fixed-effect parameter for prior attainment in Spanish Language from the full bivariate CVA model for progress in Mathematics and Spanish Language.

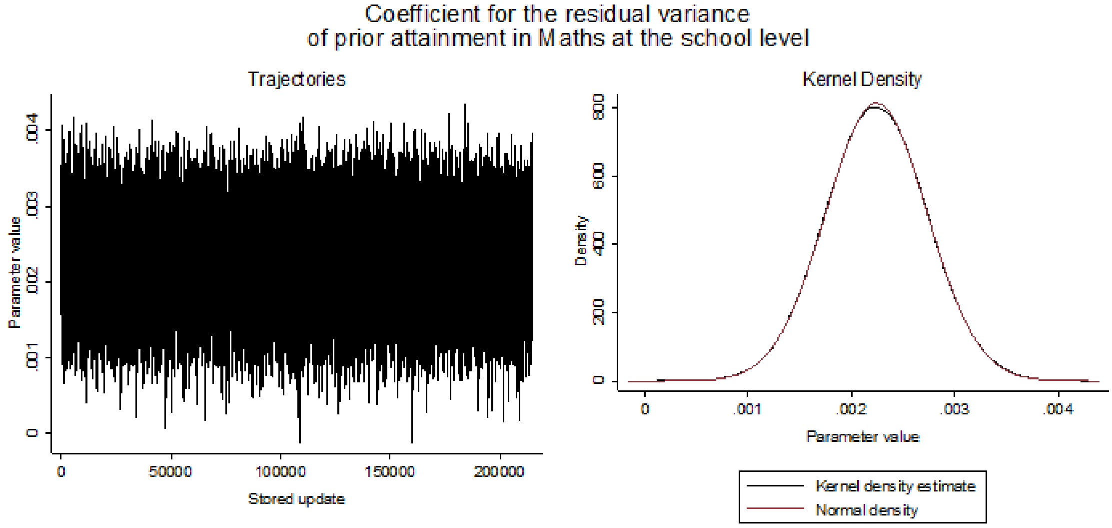


2.5. Extended MCMC information of the random-effects parameters from full bivariate 5-level CVA model for progress in Mathematics and Spanish Language.

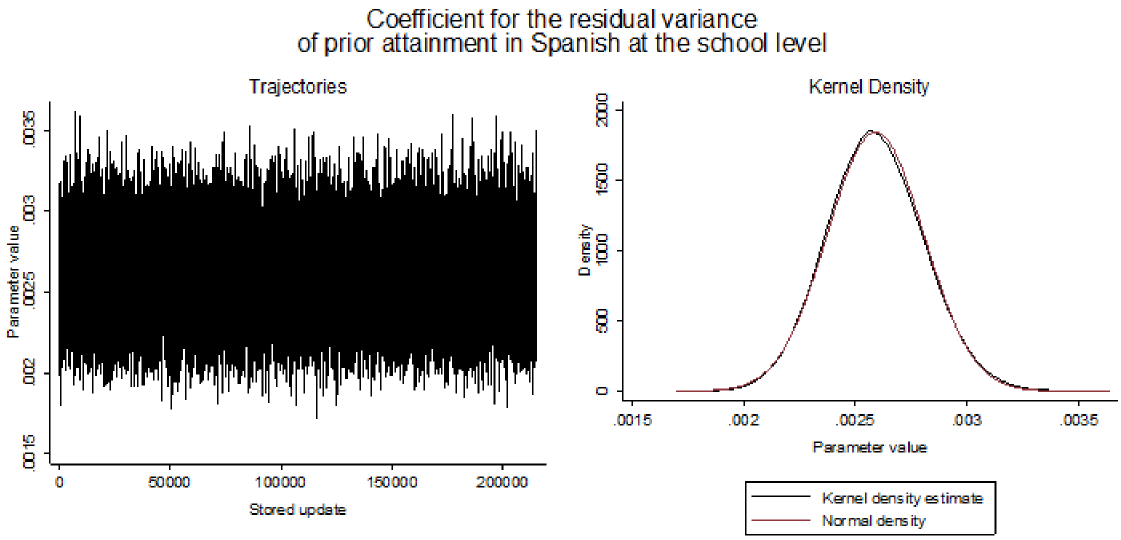
Levels	Parameters†	Coef.	S.E.	Median	Corr.	95% C.I.		ESS
Secondary Schools	Var. (Intercept Maths)	0.059	0.003	0.059	--	0.054	0.065	17160
	Var. (Intercept Language)	0.028	0.002	0.028	--	0.025	0.031	9180
	Cov. (Int. Maths, Int. Language)	0.035	0.002	0.035	0.873	0.032	0.039	15182
	Cov. (Prior att. Maths, Int. Maths)	0.003	0.001	0.003	0.204	0.002	0.004	15864
	Cov. (Prior att. Maths, Int. Lang.)	0.002	0.0005	0.002	0.214	0.001	0.003	12349
	Var. (Prior att. Maths)	0.004	0.0003	0.004	--	0.003	0.005	14733
	Cov. (Prior att. Lang., Int. Maths)	0.003	0.001	0.003	0.288	0.002	0.005	9287
	Cov. (Prior att. Lang., Int. Lang.)	0.003	0.0005	0.003	0.344	0.002	0.004	6307
	Cov. (Prior att Lang, Prior att. Maths)	0.003	0.0002	0.003	0.836	0.002	0.003	13011
	Var. (Prior att. Language)	0.002	0.0003	0.002	--	0.002	0.003	3752
	Cov. (Male Maths, Int. Maths)	0.002	0.001	0.002	0.111	-0.0004	0.004	7001
	Cov. (Male Maths, Int. Language)	0.002	0.001	0.002	0.149	0.0001	0.003	6024
	Cov. (Male Maths, Prior att. Maths)	0.00004	0.0003	0.00004	0.01	-0.001	0.001	5780
	Cov. (Male Maths, Prior att. Lang.)	-0.0001	0.0003	-0.0001	-0.044	-0.001	0.0005	2998
	Var. (Male Maths)	0.004	0.001	0.004	--	0.003	0.006	2668
	Cov. (Male Lang., Int. Maths)	0.004	0.001	0.004	0.292	0.002	0.006	3532
	Cov. (Male Lang., Int. Lang.)	0.004	0.001	0.004	0.422	0.002	0.005	2399
	Cov. (Male Lang., Prior att. Maths)	0.0002	0.0003	0.0002	0.054	-0.0005	0.001	2668
	Cov. (Male Lang., Prior att. Lang.)	0.0005	0.0003	0.0005	0.182	-0.0001	0.001	1844
	Cov. (Male Lang., Male Maths)	0.002	0.001	0.002	0.647	0.001	0.003	1598
	Var. (Male Lang.)	0.003	0.001	0.003	--	0.002	0.004	1039
Classes	Var. (Intercept Maths)	0.039	0.001	0.039	--	0.037	0.041	35076
	Cov. (Int. Maths, Int. Language)	0.025	0.001	0.025	0.784	0.023	0.026	32363
	Var. (Intercept Language)	0.025	0.001	0.025	--	0.023	0.026	22216
Pupils	Var. (Intercept Maths)	0.276	0.001	0.276	--	0.274	0.278	159869
	Cov. (Int. Maths, Int. Language)	0.079	0.001	0.079	0.248	0.077	0.081	144104
	Var. (Intercept Language)	0.368	0.001	0.368	--	0.365	0.37	157203
Local Authorities	Var. (Intercept Maths)	0.004	0.001	0.004	--	0.002	0.006	2294
	Cov. (Int. Maths, Int. Language)	0.003	0.001	0.003	0.858	0.001	0.004	2604
	Var. (Intercept Language)	0.002	0.001	0.002	--	0.001	0.004	3368
Primary Schools	Var. (Intercept Maths)	0.007	0.0004	0.007	--	0.007	0.008	14625
	Cov. (Int. Maths, Int. Language)	0.004	0.0003	0.004	0.765	0.003	0.005	8322
	Var. (Intercept Language)	0.004	0.0003	0.004	--	0.003	0.004	3487

† These parameters have been obtained via MCMC estimation with Gibbs sampling. Monitoring chain length: 215,000; burn-in: 500; storing all iterations. Trajectories are depicted for the shaded parameter values only as they are representative of all the estimated fixed-effects parameters. More details are given below. ESS=Effective sample size.

2.6. Trajectories and Density of the random-effect parameter for the coefficient of prior attainment in Mathematics at the secondary school level from the full bivariate CVA model for progress in Mathematics and Spanish Language.



2.7. Trajectories and Density of the random-effect parameter for the coefficient of prior attainment in Spanish Language at the secondary school level from the full bivariate CVA model for progress in Mathematics and Spanish Language.



Appendix 3: Extended information of full model from chapter 6

3.1. Random-effects parameters of the full bivariate model for internal school accountability and progress in Mathematics and Spanish Language.

Parameter	Estimate	S.E.	Correlation	95% Coverage int.	
Level 3: Local Authority					
Var. (Intercept Maths)	0.010	0.002	1.000	0.006	0.014
Cov. (Int. Maths, Int. Language)	0.008	0.001	0.961	0.005	0.011
Var. (Intercept Language)	0.006	0.001	1.000	0.004	0.009
Level 2: School					
Var. (Intercept Maths)	0.071	0.002		0.067	0.076
Cov. (Prior att. Maths, Int. Maths)	-0.0002	0.001	-0.008	-0.001	0.001
Var. (Prior att. Maths)	0.006	0.0003		0.005	0.006
Cov. (Int. Maths, Int. Language)	0.043	0.002	0.834	0.040	0.046
Cov. (Prior att. Maths, Int. Lang.)	-0.0001	0.001	-0.008	-0.001	0.001
Var. (Intercept Language)	0.037	0.001		0.034	0.040
Cov. (Prior att. Lang., Int. Maths)	-0.00001	0.001	-0.0003	-0.001	0.001
Cov. (Prior att. Lang., Prior att. Maths)	0.004	0.0002	0.862	0.003	0.004
Cov. (Prior att. Lang., Int. Lang.)	0.001	0.0004	0.084	0.00005	0.002
Var. (Prior att. Language)	0.003	0.0003		0.003	0.004
Level 1: Pupil					
Var. (Intercept Maths)	0.307	0.001		0.305	0.310
Cov. (Int. Maths, Int. Language)	0.096	0.001	0.280	0.094	0.098
Var. (Intercept Language)	0.383	0.001		0.381	0.386

3.2. Model fit information of the full bivariate model for internal school accountability and progress in Mathematics and Spanish Language.

Statistic†	Value
Log likelihood	-320252
Deviance	640503
AIC	640611
Number of parameters	54
Units: Local Authorities	320
Units: Schools	2438
Units: Pupils	183142

† Model estimated via the IGLS algorithm.

Appendix 4: Imputation model

4.1. Comparison between complete-case analysis and imputation model on a sample of the full dataset.

	Complete cases			Imputation model		
Fixed Part	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Pupil-level‡						
Intercept	-0.234	0.041	***	-0.234	0.039	***
Prior attainment	0.633	0.017	***	0.631	0.017	***
Male	0.060	0.029	*	0.049	0.029	
Lower-middle income	0.051	0.036		0.051	0.039	
Upper-middle income	0.059	0.055		0.045	0.056	
High income	0.084	0.088		0.086	0.091	
Held back	-0.176	0.049	***	-0.188	0.049	***
School-level§						
Subsidised independent	0.052	0.035		0.052	0.033	
Independent	0.089	0.158		0.098	0.155	
Lower-middle SES	0.090	0.044	*	0.084	0.043	*
Middle SES	0.262	0.054	***	0.273	0.052	***
Upper-middle SES	0.461	0.072	***	0.483	0.071	***
Upper SES	0.521	0.178	**	0.518	0.175	**
Random Part	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Level 2: School						
Var (Intercept)	0.073	0.016		0.055	0.015	
Level 1: Pupil						
Var (Intercept)	0.284	0.017		0.314	0.017	
Model information						
Deviance	3127.1			N/A		
Units: Schools	1277			1277		
Units: Pupils	1743			1914		

* p<0.05 ** p<0.01 *** p<0.001

‡ Reference categories for pupil-level variables: Female; Low income and Not held back.

§ Reference categories for school-level variables: State-funded school and Low school SES.

The table above compares the results of two simple 2-level CVA models fitted on a 1% sample of the full dataset. The first model is a complete-case analysis, in which missing cases are deleted listwise. The second model corresponds to the pooled results of a model fitted on 10 imputed datasets. The imputation model was run using the software package REALCOM-IMPUTE (Carpenter et al., 2011). The model of interest and the imputation model were kept deliberately simple, because of the computational burden of complex models.

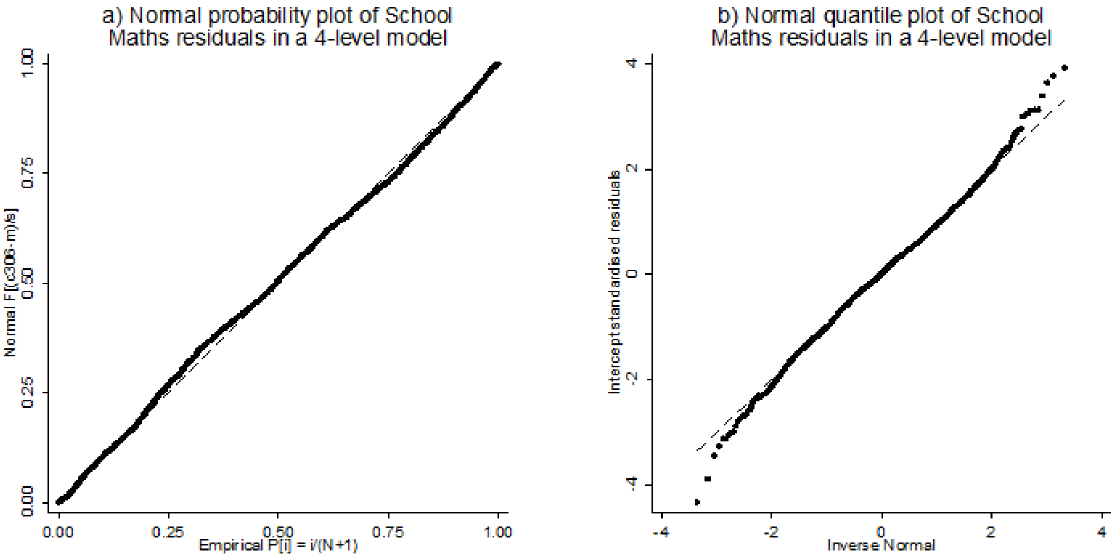
The estimates show that very little is found the fixed-effects coefficients. With the sole exception of the coefficient for male pupils, the rest of the coefficients have remained unchanged. The imputation model shows that the effect of being male is not significant when using the multiply-imputed datasets; however, it is known from the models presented in chapters 4, 5 and 6 that this effect significantly interacts with other variables at the level of pupils, i.e. prior attainment, year repetition and income. The omission of those interactions might well explain the change in the coefficient.

With respect to the random-effects parameters, it is appreciated that the pooled results show a different variance partitioning between pupils and schools. This is by no means a concern, insofar as the results from the imputation model constitute further evidence of the overestimation of the school effects in the traditional models. Furthermore, the increased variance at the pupil level in the pooled results might well be due to the omission of the classroom level.

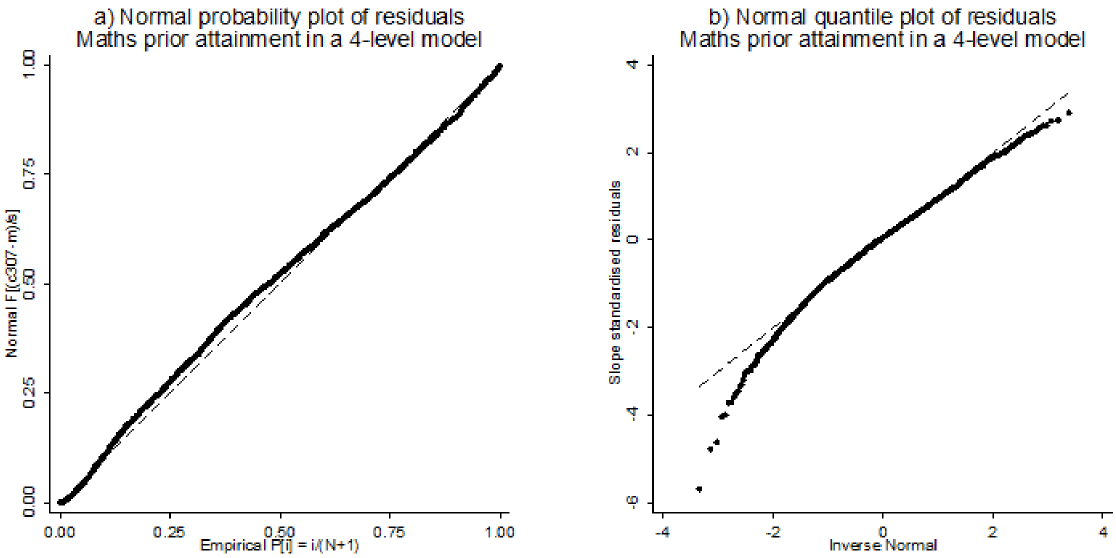
This imputation analysis is meant to be an illustration of the potential bias in the models. In any case, as mentioned in chapter 3, the main aim of this thesis was to estimate more reliable school accountability measures, i.e. the residuals at the school level derived from the multilevel models. The currently available imputation methods are not capable of estimating these residuals on multiply-imputed datasets, and thus, this exercise can only be taken as a possible indication of how the estimates could change. Further research could explore these issues more deeply.

Appendix 5: Assumptions checking

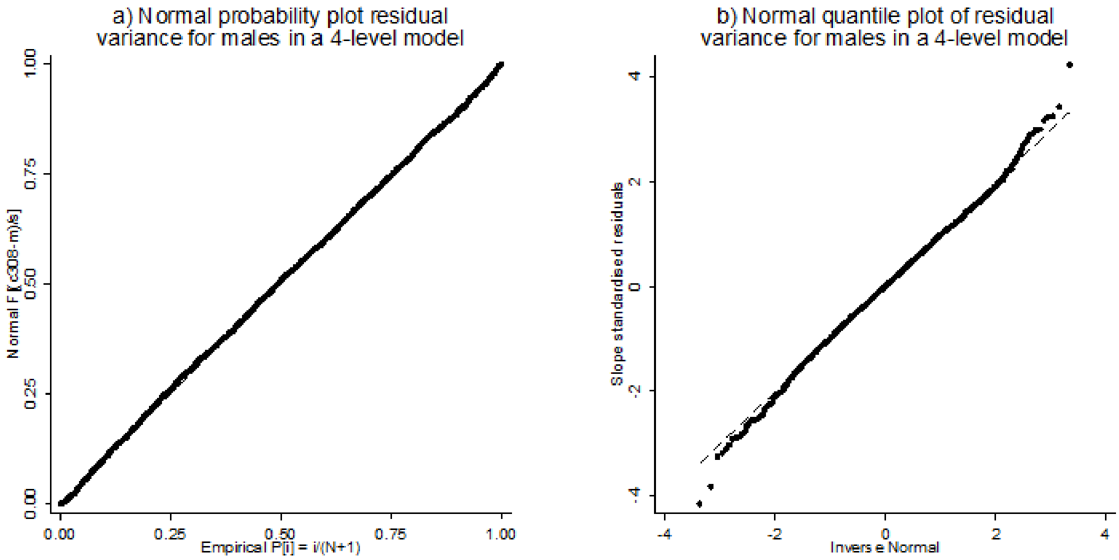
5.1. Normality of intercept residuals at the secondary school level from full 4-level CVA model for progress in Mathematics (Chapter 4, Tables 4.12 and 4.13)



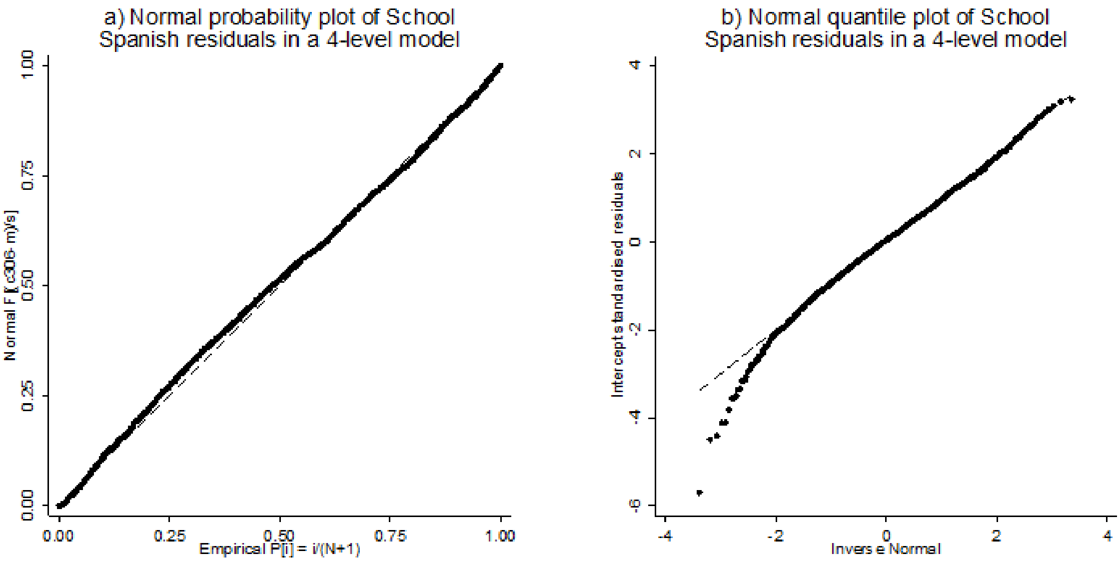
5.2. Normality of prior attainment slope variance residuals at the secondary school level from full 4-level CVA model for progress in Mathematics (Chapter 4, Tables 4.12 and 4.13)



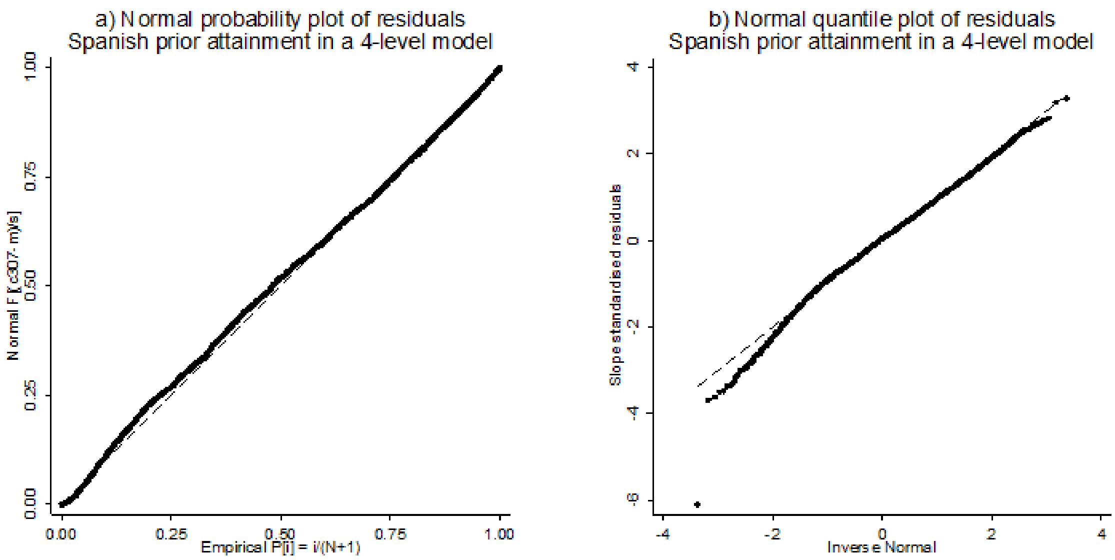
5.4. Normality of male pupils' variance residuals at the secondary school level from full 4-level CVA model for progress in Mathematics (Chapter 4, Tables 4.12 and 4.13)



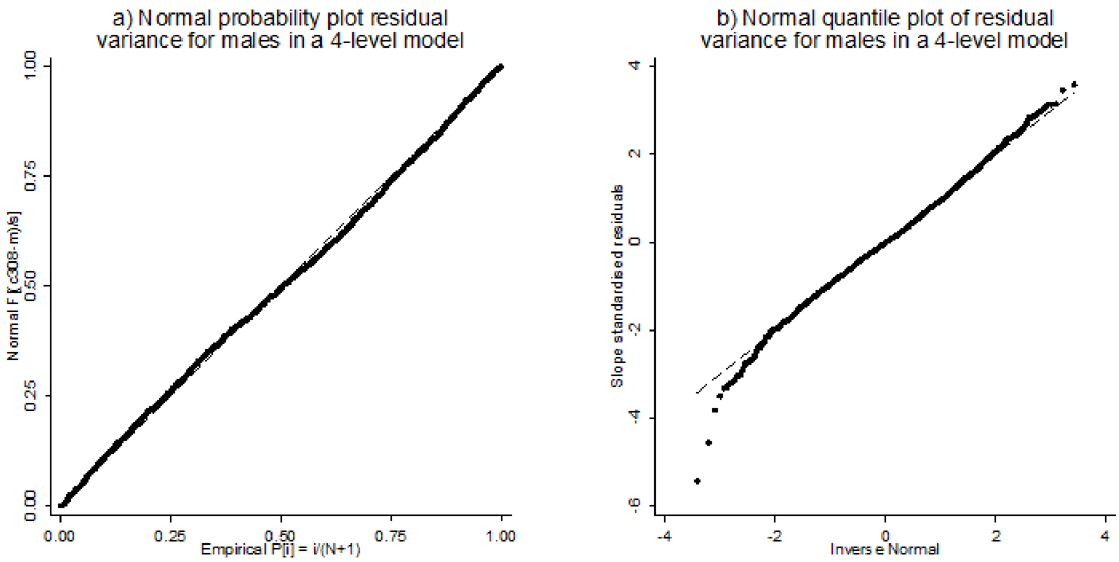
5.4. Normality of intercept residuals at the secondary school level from full 4-level CVA model for progress in Spanish Language (Chapter 4, Tables 4.17 and 4.18)



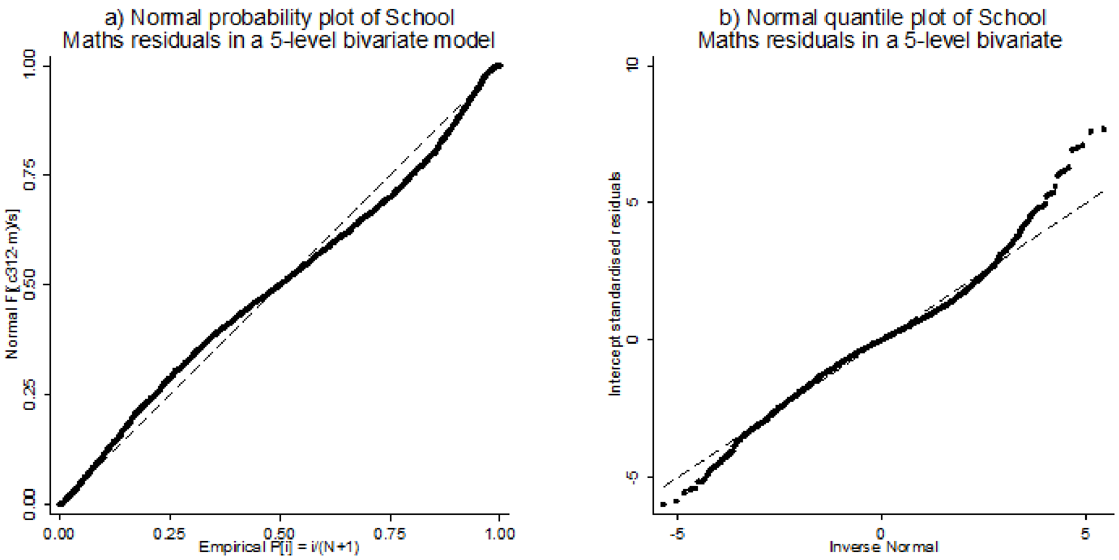
5.5. Normality of prior attainment slope variance residuals at the secondary school level from full 4-level CVA model for progress in Spanish Language (Chapter 4, Tables 4.17 and 4.18)



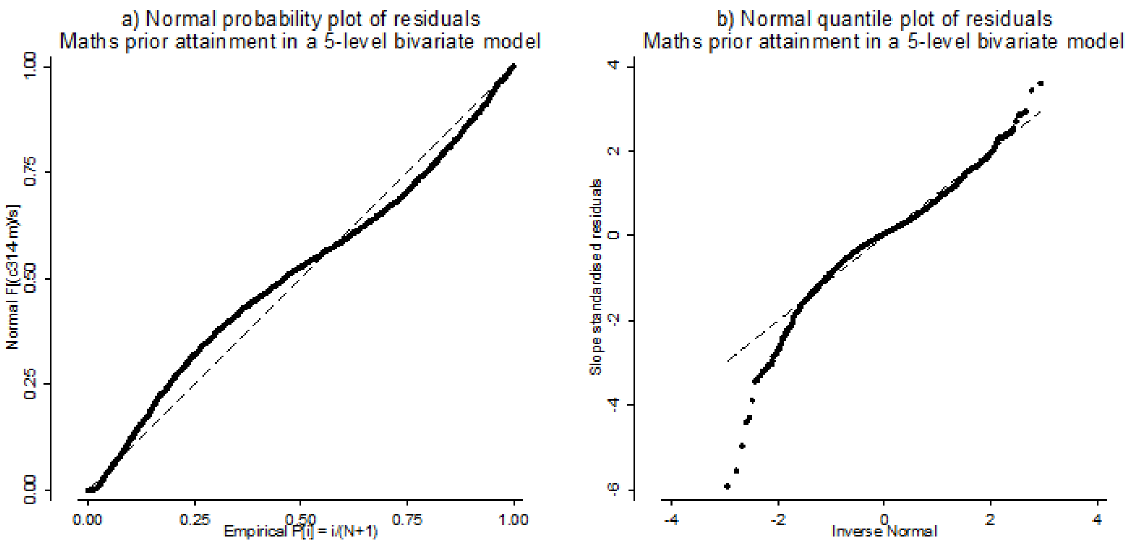
5.6. Normality of male pupils' variance residuals at the secondary school level from full 4-level CVA model for progress in Spanish Language (Chapter 4, Tables 4.17 and 4.18)



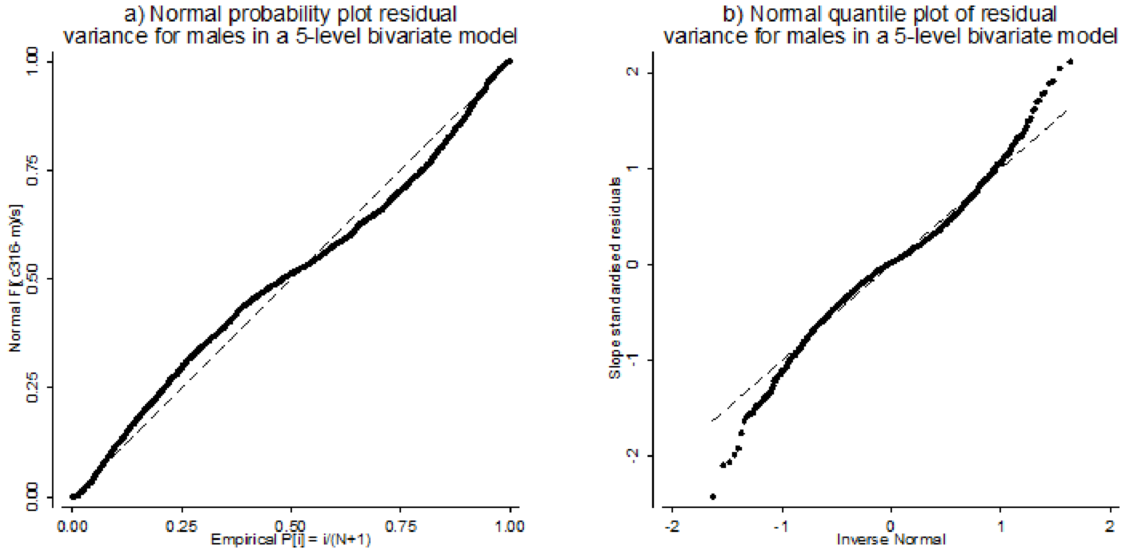
5.7. Normality of intercept residuals at the secondary school level in Mathematics from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



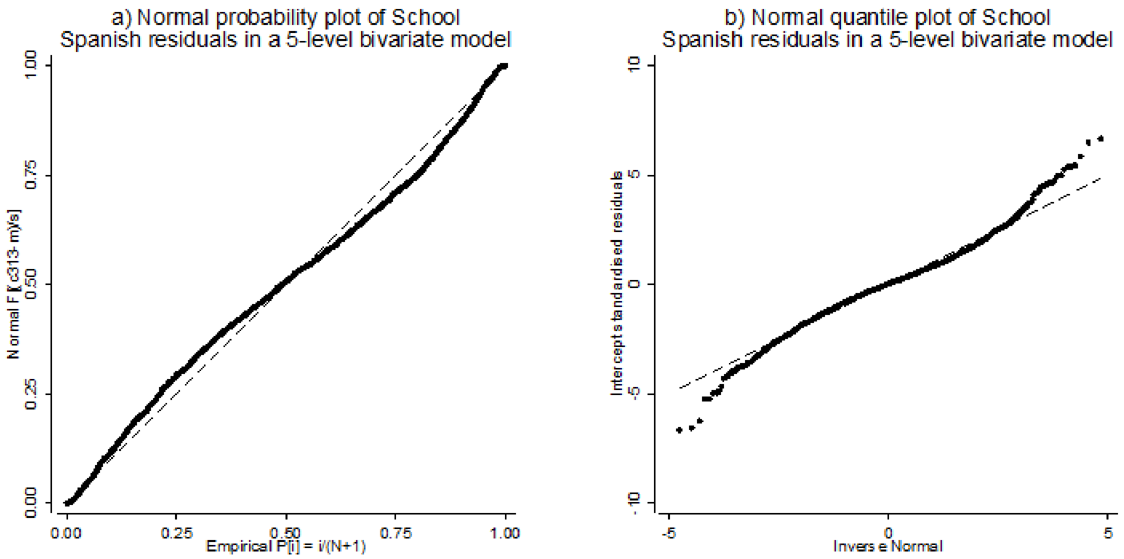
5.8. Normality of prior attainment slope variance at the secondary school level in Mathematics from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



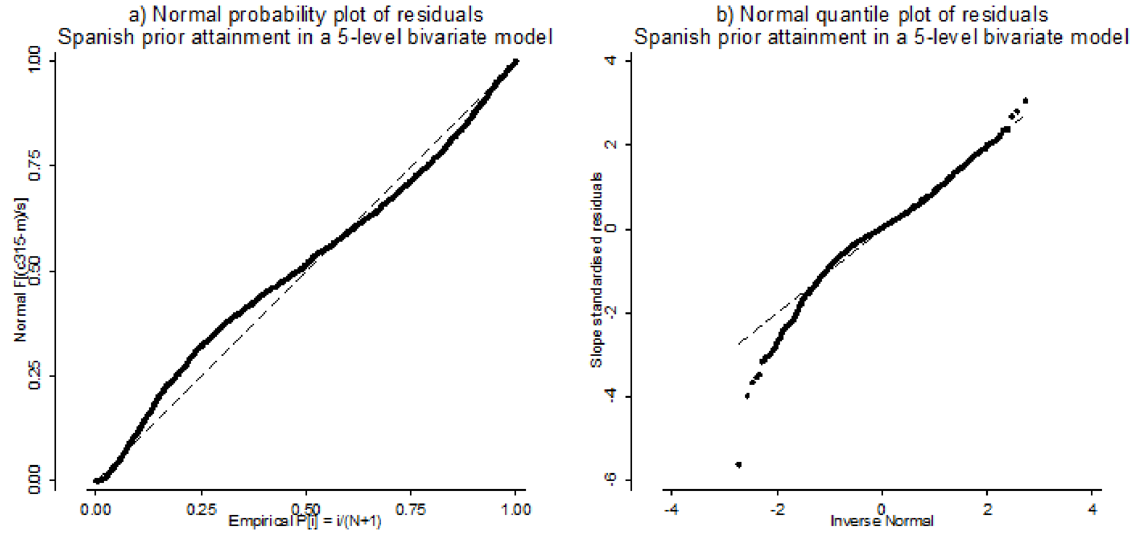
5.9. Normality of male pupils' variance residuals at the secondary school level in Mathematics from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



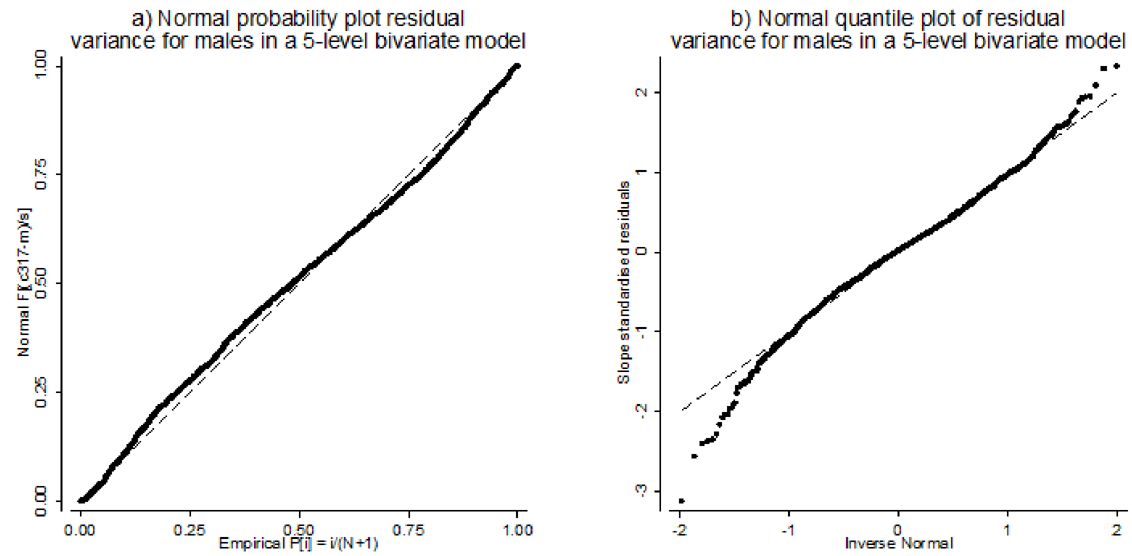
5.10. Normality of intercept residuals at the secondary school level in Spanish Language from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



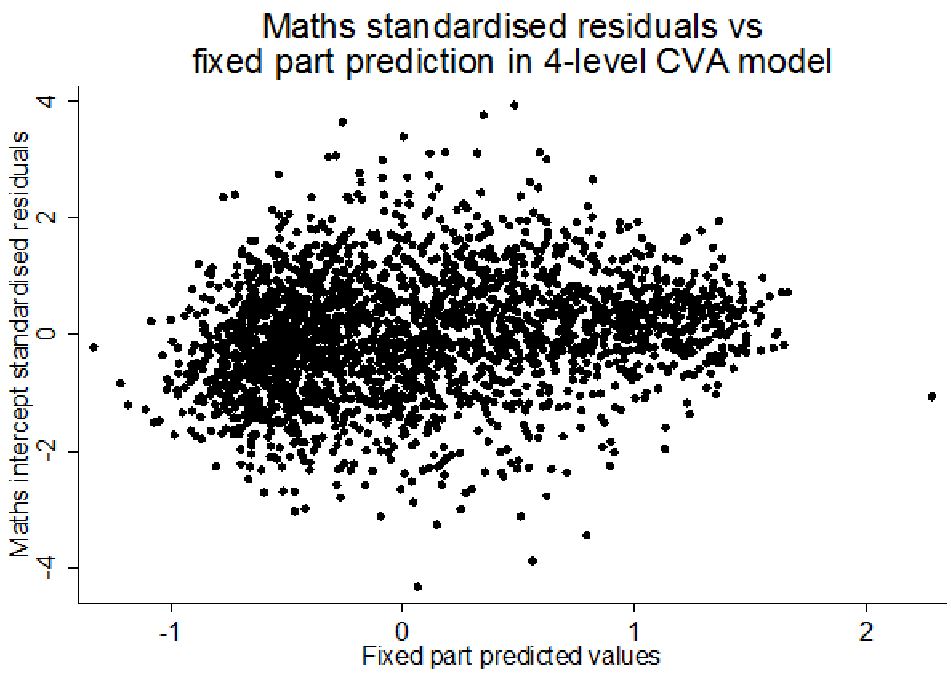
5.11. Normality of prior attainment slope variance at the secondary school level in Spanish Language from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



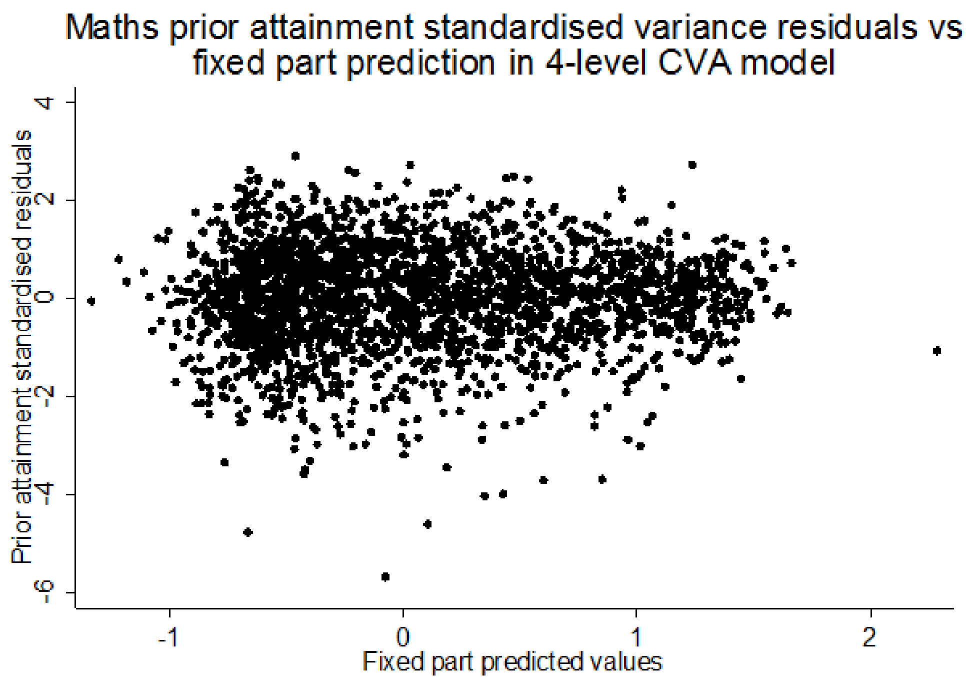
5.12. Normality of male pupils' variance residuals at the secondary school level in Spanish from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



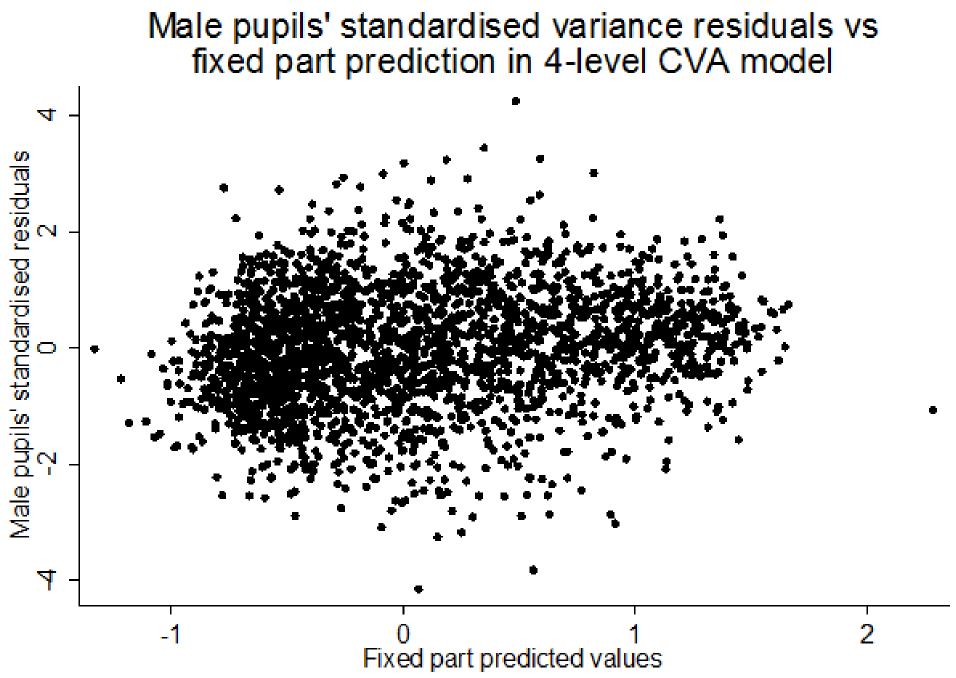
5.14. Homoscedasticity of intercept residuals at the secondary school level from full 4-level CVA model for progress in Mathematics (Chapter 4, Tables 4.12 and 4.13)



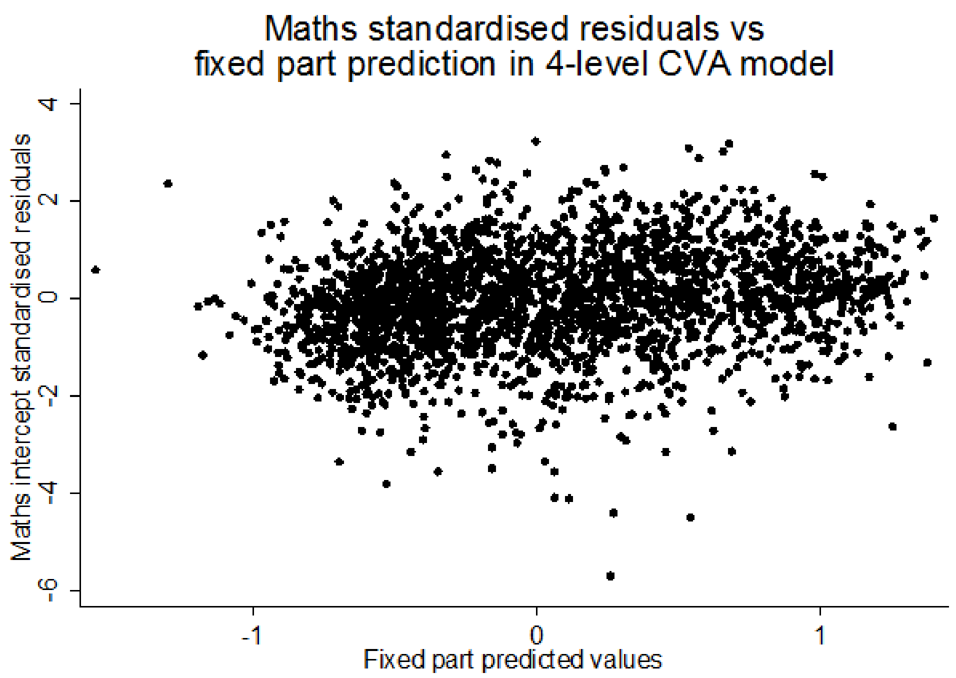
5.14. Homoscedasticity of prior attainment slope variance residuals at the secondary school level from full 4-level CVA model for progress in Mathematics (Chapter 4, Tables 4.12 and 4.13)



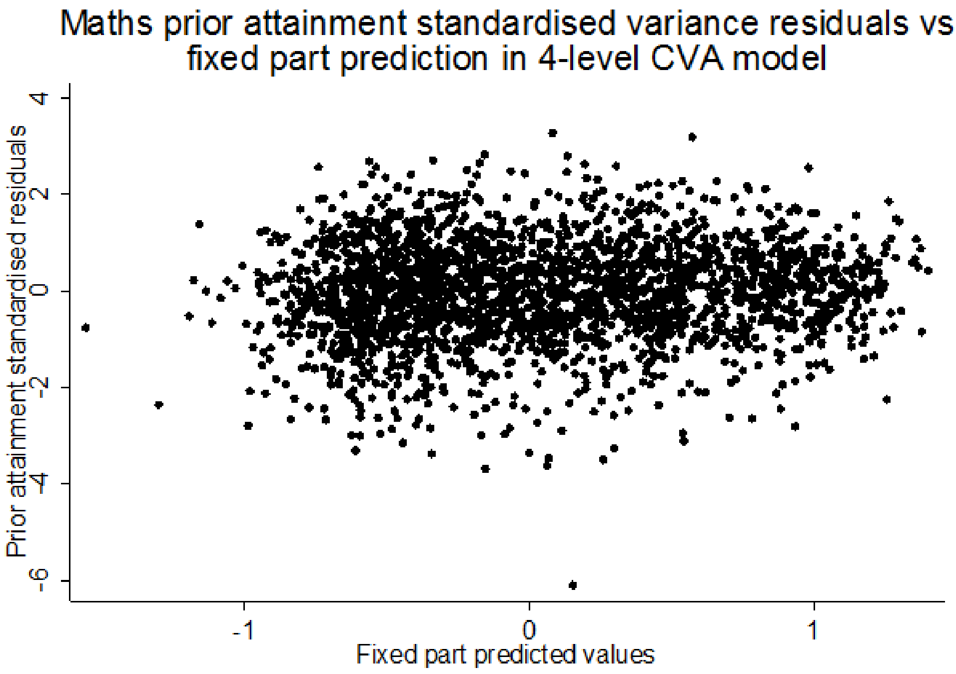
5.15. Homoscedasticity of male pupils' variance residuals at the secondary school level from full 4-level CVA model for progress in Mathematics (Chapter 4, Tables 4.12 and 4.13)



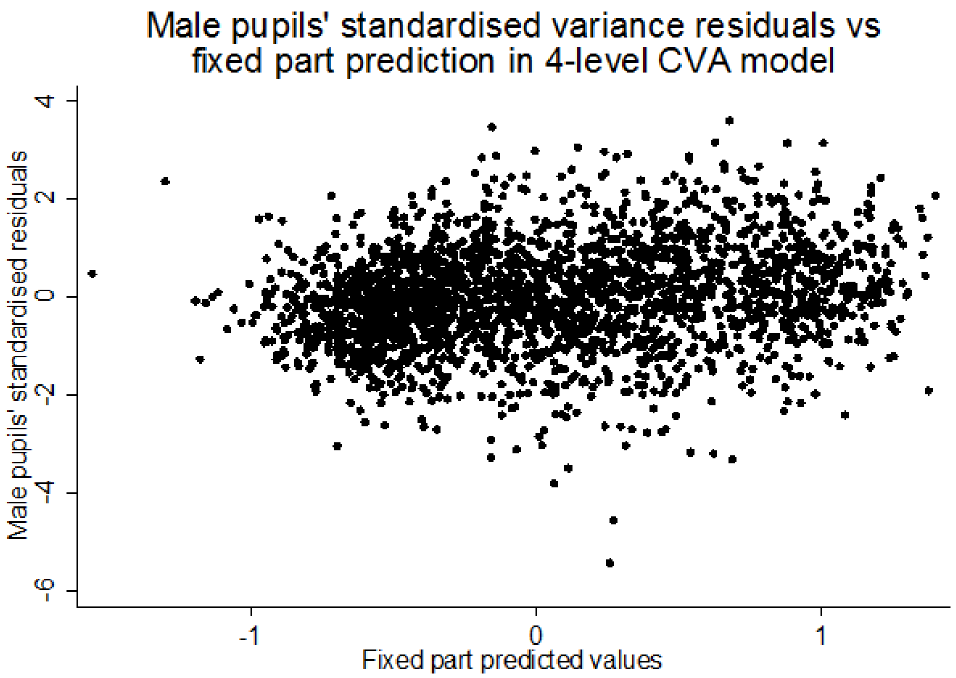
5.16. Homoscedasticity of intercept residuals at the secondary school level from full 4-level CVA model for progress in Spanish Language (Chapter 4, Tables 4.17 and 4.18)



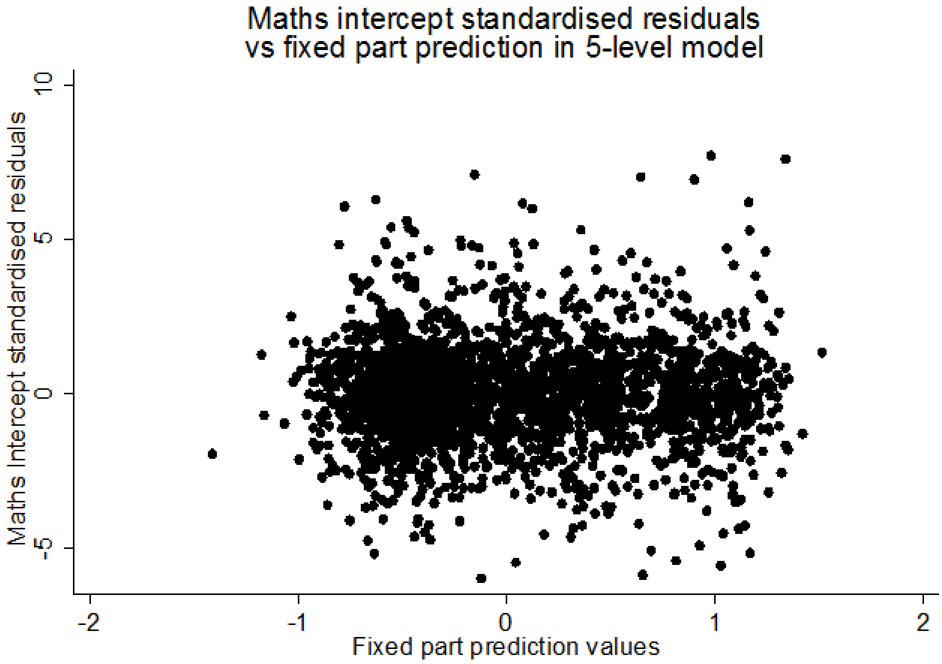
5.17. Homoscedasticity of prior attainment slope variance residuals at the secondary school level from full 4-level CVA model for progress in Spanish Language (Chapter 4, Tables 4.17 and 4.18)



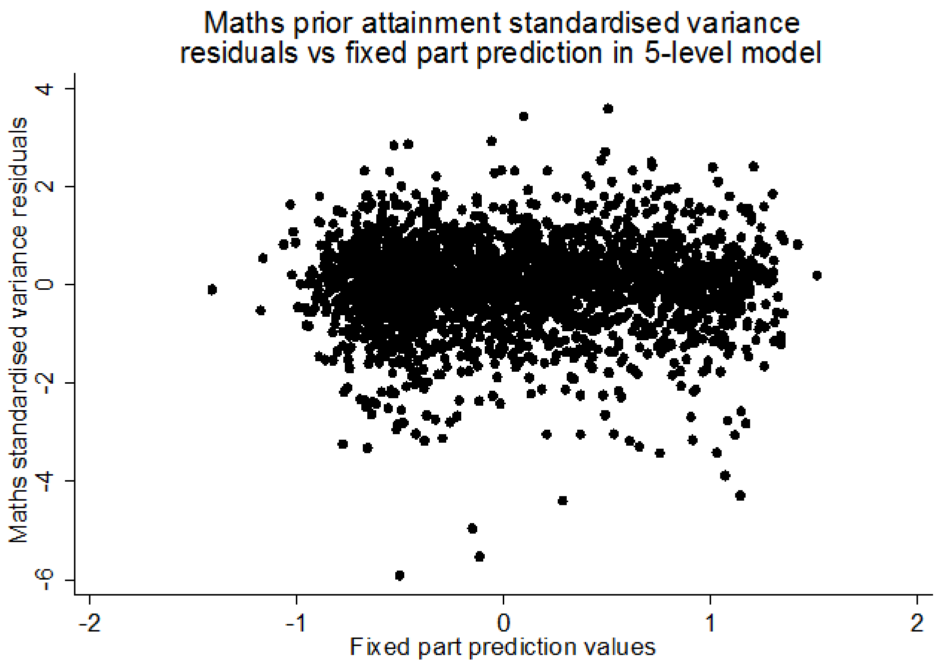
5.18. Homoscedasticity of male pupils' variance residuals at the secondary school level from full 4-level CVA model for progress in Spanish Language (Chapter 4, Tables 4.17 and 4.18)



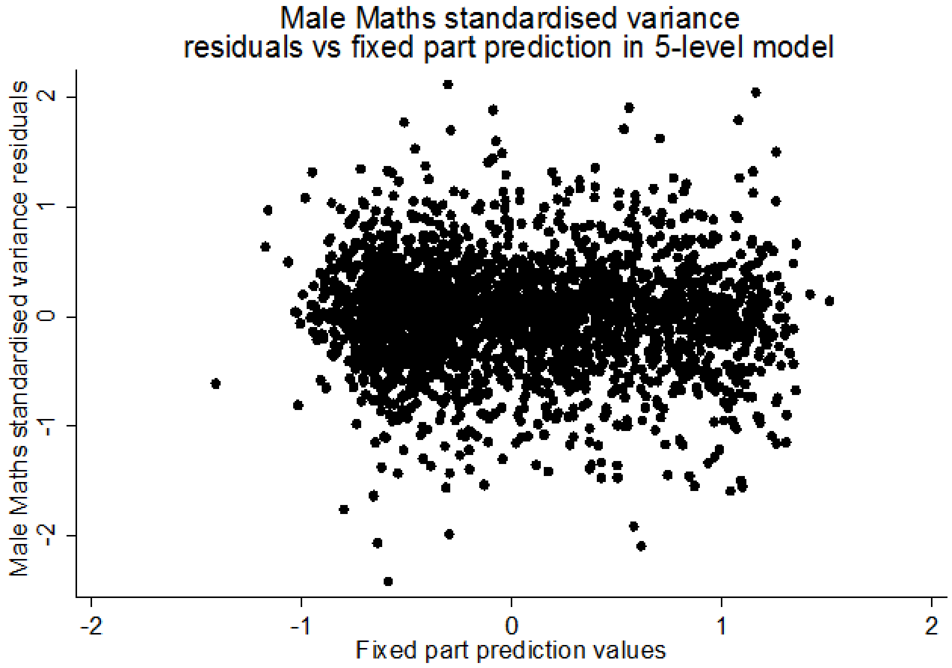
5.19. Homoscedasticity of intercept residuals at the secondary school level in Mathematics from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



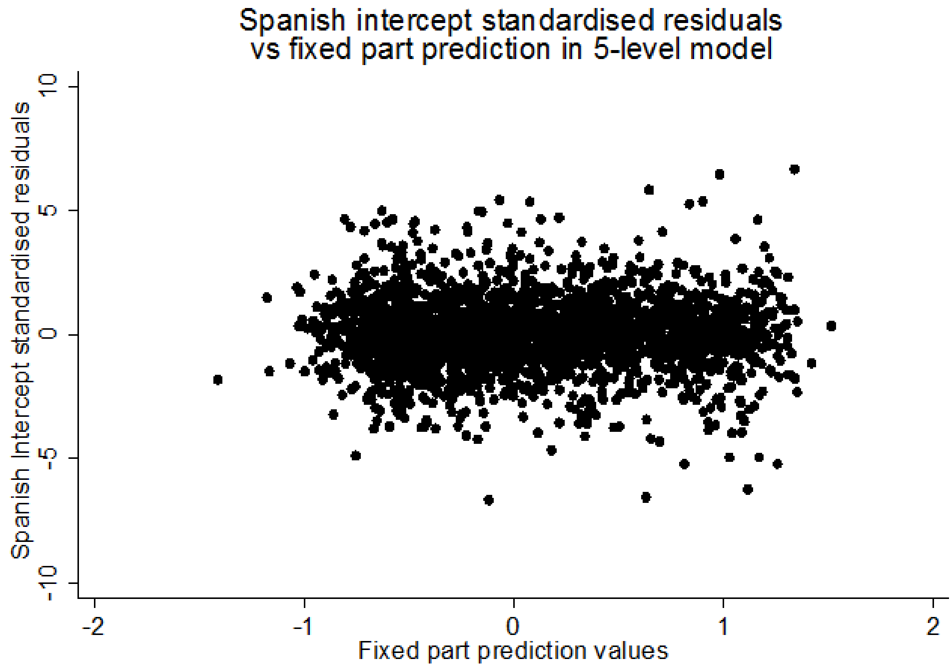
5.20. Homoscedasticity of prior attainment slope variance at the secondary school level in Mathematics from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



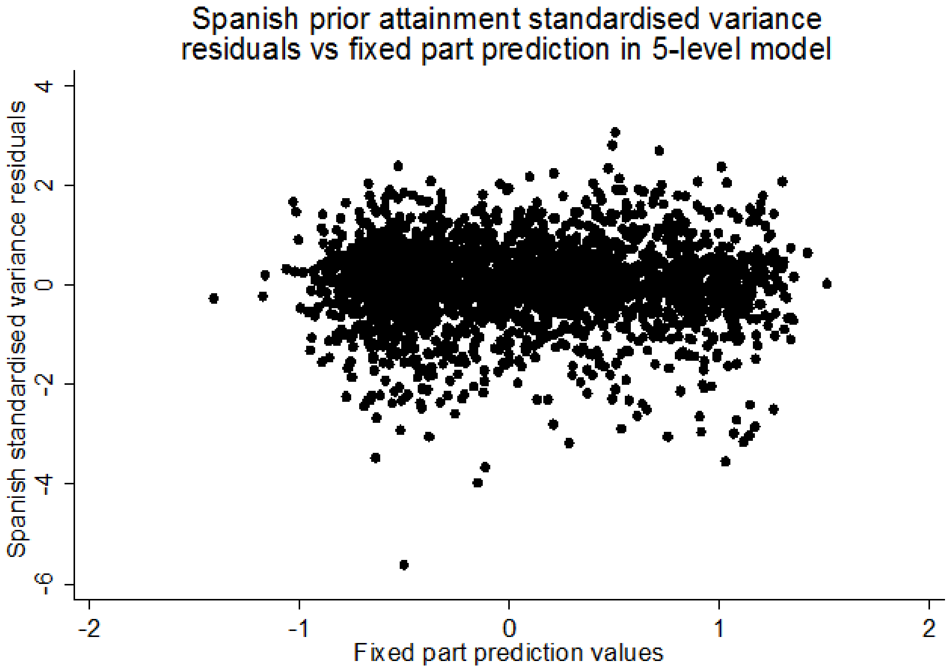
5.21. Homoscedasticity of male pupils' variance residuals at the secondary school level in Mathematics from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



5.22. Homoscedasticity of intercept residuals at the secondary school level in Spanish Language from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



5.23. Homoscedasticity of prior attainment slope variance at the secondary school level in Spanish Language from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)



5.24. Homoscedasticity of male pupils' variance residuals at the secondary school level in Spanish from full 5-level bivariate CVA model for progress in Mathematics and Language (Chapter 5, Tables 5.15 and 5.16)