Educational Effectiveness in Chilean secondary education: comparing different 'value added' approaches to evaluate schools

B. Muñoz-Chereau, Graduate School of Education, University of Bristol, Bristol, England

S.M Thomas, Graduate School of Education, University of Bristol, Bristol, England.

Abstract

This article reports an original investigation into school performance measures and the multilevel nature of pupil achievement data in the Chilean school system using a sample of 177,461 students, nested within 7,146 classrooms, 2,283 secondary schools and 313 municipalities. The dataset comprised Year 10 Students' 2006 SIMCE tests results in two subject outcomes (language and maths) matched to their prior attainment in grade 8 and family characteristics. The analyses showed the lack of precision of two level models to draw conclusions regarding the effectiveness of Chilean secondary schools as well as the extent to which different pupil intake, background and context features of Chilean secondary schools influence students' performance. The results show substantial and statistically significant municipal, school and classroom effects in Chile and how the estimation of school effects changes according to the explanatory variables controlled for in the analysis and the outcome analysed. These results are compared with similar studies carried out in Latin-America, as well as in other countries (England and China), in order to situate the findings in the broader knowledge-base of Educational Assessment and Effectiveness Research.

Keywords: value added; school effectiveness; educational quality; mathematics achievement; multilevel analysis; Chilean secondary education

Introduction

The knowledge-base of Educational Effectiveness Research has indicated that the assessment and estimation of school effects using student achievement or examination data can be strongly affected by various factors. These include not just the type of explanatory variables adjusted for in the analyses (eg prior achievement, pupil background characteristics, school context), the outcomes employed (eg language or mathematics) or the student group sampled (Sammons, Thomas and Mortimore, 1997), but also crucially the hierarchical 'clustering' levels specified in a model. Previous research mostly conducted in Europe and USA has demonstrated that the structural levels in the education system below (classroom, department) and above (municipality, province, country, etc.) the school level, have played an important role in different contexts when exploring the relative importance of schools (Bosker and Scheerens, 1989; Goldstein, 1997; Opdenakker and Van Damme, 2000; Thomas, 2001; Luyten, 2003; Van den Noortgate, Opdenakker and Onghena, 2005; Cervini, 2009; Martínez, 2012). By not specifying levels both below and above the school in multilevel analyses when estimating school effects, model results can potentially be both miss-specified and misleading in terms of overestimating school effects. However, this aspect is often not highlighted, due in part to data limitations in national datasets that may not include unique class and regional identifiers as well as student and school identifiers.

In spite of this limitation the evaluation of school effects based on detailed analyses of student assessment and attainment data is increasingly being used as a key strand of school inspection and self evaluation systems across the world, so it is essential to explore the relative importance of each level of an education system as a potential source of variation in student outcomes. This issue is especially pertinent in the Chilean context where school performance measures are soon to be implemented as part of a new high stakes accountability system. Thus this study has two key aims: to investigate the precision of two, three and four level models to draw conclusions regarding the effectiveness of Chilean secondary schools and also the extent to which different pupil intake, background and context features of Chilean secondary schools influence students' performance.

The study is very timely given that in recent years Chile has given way to a major policy shift towards the introduction of new laws and accountability mechanisms (Treviño and Donoso, 2010). The approval by the Congress of the Ley General de Educación (LGE) put in place in 2013 the 'Sistema Nacional de Aseguramiento de la Calidad de la Educación' (National System of Assurance of the Quality in Education) that encompasses the creation of two new bodies: the Superintendence of Education and the Quality Agency of Education (QAE). Since last year the QAE – a local version of the Office for Standards in Education (OFSTED) – is mandated to not just classify schools in four performance groups (Good, Satisfactory, Fair and Poor) based on school SIMCE assessment results and other data, but also will be implementing high stakes consequences on the basis of this classification (ranging from organisational interventions up to school closure) for those schools that do not meet the required standards. However, the well-known reality is that Chile has a stronger than average relationship between student academic performance and socio-economic background[1]. Thus, it could be seen as highly problematic that the new institutional assessment framework developed by the QAE does not intend to use more sophisticated and sensitive approaches to evaluating school performance that could statistically adjust for these factors, such as 'contextualised value added' measures (QAE, 2013). The implication is that the annual categorisation of schools will not take proper account of contextual or compositional effects at the level of the school, suggesting that the collective effects of group processes on individuals' performances are not important; nor will the influence of other 'clustering' levels within the education system (such as classes or municipalities) be considered.

Therefore this paper seeks to review pertinent studies carried in Chile and elsewhere, as well as build on this work by contributing further relevant evidence to the on-going debate about fairer ways to judge the effectiveness of a school (Thomas and Mortimore, 1996; Creemers; 1997; Saunders, 2000; Ferrer, 2006; OECD, 2008). New evidence is crucial to inform and potentially improve the future credibility and legitimacy of the school evaluation and assessment systems under implementation in Chile and in other country contexts.

Literature review

Measures of School performance and Effectiveness: Raw, Contextualized Attainment (CA), Value Added (VA) and Contextualised Value Added (CVA)

Raw attainment measures

In the international scenario there is considerable literature from western countries in general (Willms & Raudenbush, 1989; Raudenbush & Willms, 1995; Ladd & Zelli, 2002; Hoyle & Robinson, 2003; OECD, 2008, 2011a; Rosenkvist, 2010; Scherrer, 2011) and the UK in particular (Goldstein et al., 1993; Sammons et al., 1997; Stoll & Mortimore, 1997; Slee, Weiner & Tomlinson, 1998; Sammons, 1999; De Luca, in Lambert & Lines, 2000; Leckie & Goldstein, 2007; Leckie, 2008; Thomas, Salim, Muñoz-Chereau, & Peng, 2012) supporting the argument that raw or unadjusted measures of pupil achievement in a given examination or test provide a poor method for comparing schools' performance. Particularly in Chile, the decision to judge school performance based on their raw test results would be a way of rewarding social segregation (Treviño & Donoso, 2010). Even though concluding from raw results that schools are effective or ineffective would be misguided, the important interpretation that can be derived is the considerable size of the achievement gap between schools in many educational systems. This points to the necessity to put raw results into context, and has supported the development of better measures for judging school performance. The claim

1 While the average percentage of total variance explained by socioeconomic background factors across OECD countries is 14%, in Chile the equivalent figure is 57% (OECD, 2010).

that something more sophisticated than raw 'league tables' is needed in order to compare schools on a more equitable basis has been strongly associated with VA and CVA approaches (Schagen & Hutchison, 2003), because they offer the promise of a more rigorous approach for levelling the playing field (Braun, Chudowsky & Koenig, 2010; Scherrer, 2011).

Raw attainment studies carried out in Latin-America and the Caribbean region

Although Latin-America and the Caribbean educational systems are strongly associated with pressing but unfulfilled quality and equity issues in student performance, not all is discouraging in the region. Hanushek, Peterson and Woessman (2012) compared the rates of achievement growth in math, science and reading in 49 countries based on 28 administrations of PISA, TIMSS and PIRLS implemented between 1995 and 2009. They highlighted that Brazil, Chile and Colombia were among the ten countries with the largest gains in average student performance, estimated as a test-score gain from year to year (percent of standard deviation). Gamboa and Waltenberg (2012) analysed PISA 2006–2009 results in mathematics, reading, and sciences in terms of the extent to which six Latin-American countries (Argentina, Brazil, Chile, Colombia, Mexico and Uruguay) educational achievements are explained by what they called 'pupils' circumstances', such as gender, parental education or school type. They concluded that inequality of opportunity is substantially oscillating in the region. In terms of country differences 'Brazil stands out as the country presenting more inequality of opportunity in 2006, followed by Mexico; the countries presenting less inequality are Argentina and Colombia. In 2009, unfairness is higher in Uruguay and Brazil, and lower in Mexico and Chile' (p.695-6). Moreover, at the primary level the Second Regional Comparative and Explanatory Study (SERCE-UNESCO, 2008) also gave insights into the learning achievements of a representative Latin-American and Caribbean sample of 100,752 Third Grade Students and 95,288 Sixth Grade students nested within 3,065 schools in 16 countries in the subjects of Mathematics, Language (Reading and Writing) and Natural Science.

(Table 1)

Looking at Table 1, three groups of countries were identified: (1) Countries that systematically obtained scores significantly higher than the regional average (Chile, Costa Rica, Cuba, Mexico and Uruguay); (2) Countries that systematically obtained scores significantly lower than the regional average (Dominic Republic, Ecuador, Guatemala, Nicaragua, Panama and Paraguay) and (3) countries matching the regional average or exhibiting less systematic results (Argentina, Brazil, El Salvador, Colombia and Peru). In other words, to talk about 'the' Latin-American educational system would be a mistake, as substantial differences in the quality and equity of student learning across the region at both primary and secondary levels are evident.

Contextualized Attainment (CA) studies carried out in Latin-America and the Caribbean region

Differences across countries within the Latin-American region have also been highlighted by CA studies oriented to explore the multilevel nature of the educational system over the last decade. The growing availability of statistical models and theoretical consensus in favour of using MLM (Goldstein, 1995), along with the availability of powerful software to carry out this type of analysis, has started to be matched with analytic expertise and research seeking to explore levels below and beyond the school in the Latin-American region. The Second Regional Comparative and Explanatory Study (SERCE-UNESCO, 2008) mentioned above has also been utilized to estimate primary school effects in terms of 'Contextualized Attainment' due to the collection of contextual, socio-demographic, family and personal data. This information, also matched to school processes data, were collected by administering parent, student, teacher and Head-teachers' questionnaires, in addition to test results.

Extending the raw performance findings reported in Table 1, a Multilevel analysis of SERCE datasets was conducted to estimate the student outcome variability of both 'contextualized attainment' and 'raw' performance in each country/region that is attributable to differences between primary schools (Murillo and Roman, 2011). The authors analysed a randomly stratified sample of two cohorts: 90,300 Year 3 students and 86,362 Year 6 students nested within the same 2,809 schools within 15 countries (excluding Mexico because it didn't apply the context questionnaires). Using a 3-level 'clustering' model (students nested within schools within countries), they reported that for 'raw' performance measures the country effect was very significant, with estimates of the percentages of variance in student outcomes attributable to differences between countries ranging from 23% in Math Year 3 to 18% in Language Year 6 (calculated using the null model which adjust for no explanatory variables). Looking at the countries individually, they found that they varied enormously, ranging from a 'raw' primary school effect of 54% in Cuba (Math Year 6), to 6.5% in Costa Rica (Math Year 3). The authors reported results organised in three groups: Argentina, Colombia, Ecuador, Panamá and Paraguay had school effects higher than the regional mean of 20%. Brazil, Guatemala and Peru had school effects similar to the regional mean, but Costa Rica, Chile, El Salvador, Nicaragua, Dominican Republic and Uruguay, had school effects lower than 15%. Cuba was an outlier with a school effect around 46% in both subjects, which is interesting because it is the country in the region with the strongest public education system, and the only one that has not implemented neoliberal market oriented reforms. Subsequently, a Contextualised Attainment (CA) model was applied to the same dataset and provided the best goodness of fit for the data by controlling for contextual factors at the family (parents' education and family SES), school (school SES) and individual level (gender, student first language, and years of preschool). Using this approach the between country effect was slightly reduced - ranging from 20% in Math Year 3 to 15% in Language Year 6 of the total variance attributable to differences between countries. Looking at the within country between school effect for the CA model, this was also reduced to 22% in Math Year 6 and to 17.6% in Language Year 3. So Murillo and Roman (2011) found bigger country and primary school effects in math than in reading, and in Year 6 than Year 3, supporting previous similar research conducted in UK (Sammons et al., 2008). Focusing only on the Chilean sample of 6,136 students nested within 165 schools, they concluded that the between school variation in math outcome explained by student background factors alone varied from 13.7% in Year 3 to 16.1% in Year 6, whereas in reading the equivalent figures were 9.7% in Year 3 and 10.26% in Year 6. However, Murillo and Roman (2011) results need to be interpreted with caution given their methodological limitations. Firstly, the data was cross-sectional, so conclusions regarding the effectiveness of primary schools are limited given the lack of prior attainment data. Secondly, the goal of universal primary education for all children has been met in Chile, but not in the whole Latin-American region. Thirdly, and more critically, the authors did not report the standard errors of their estimates, nor the goodness of fit of their models, that is, the amount of total variance explained by the models. Despite these limitations, this study is pioneering in the examination of country and school performance effects in an international comparative context, describing a pattern of similarities and differences between primary schools in 15 Latin-American and Caribbean countries.

In line with Murillo and Roman (2011), other CA multilevel analyses carried out in the region have also highlighted the relevance of considering the hierarchical structure of the educational system when exploring school effects, because the distribution of pupils' achievement seems to vary significantly not just between pupils and schools, but also among classrooms, municipalities and states in Mexico (Blanco, 2008) and Argentina (Cervini, 2009). However, this approach has been scant in Chile. One previous study exploring the importance of different levels conducted by Ramírez (2007) did not use MLM - which would have enabled the synchronicity of levels to be analysed in the presence of clustering - but one way ANOVA, considered from a statistical point of view an inferior method that can lead to incorrect inferences (Rasbash, Jones, Steele, & Pillinger, 2009; Marsh, Lüdtke, Nagengast, Trautwein, Morin, Abduljabbar, & Köller, 2012).

Value Added (VA) Measures

Creemers (1996) among others defined VA measures as a way of estimating the quality of the school by considering the average score in a given measurement after correcting for input characteristics. The so-called VA approaches try to identify the individual school's contribution to students' relative progress over time (Aitkin & Longford, 1986; Gray, Reynolds, Fitz-Gibbon, & Jesson, 1996, Creemers, 1996; Gray, Goldstein & Thomas, 2003). This is the reason why the availability of baseline prior attainment as well as student background and other factors is crucial in estimating schools value added performance. Using this approach it is argued that 'the schools could see to what extent they have boosted pupils' progress' (Stoll & Mortimore, 1997, p.16). A school that adds value has been defined as one in which pupils progress further than might be expected from consideration of the intake (Mortimore, 1991; Sammons, Cuttance, Nuttall, & Thomas, 1995). Other authors have taken a more critical view to VA approaches by describing them as necessary to put the raw school examination results in context (Thomas & Mortimore, 1996), while others have defined them as 'methods that account for the differences in student intake among schools' (Timmermans, Doolaard, & De Wolf, 2011, p. 393). Teddlie and Reynolds (2000) defined VA as 'a measure of the relative gain in achievement made by pupils. The rationale is that a school is not responsible for the absolute level of student achievement so much as for the progress made by pupils in its care' (p. 264).

Value-added studies carried out in Latin-America and the Caribbean region

Although some researchers have studied the school effect using raw and contextualised attainment multilevel models in cross-section ways in the region - especially in Colombia (Casas, Gamboa, and Piñeros, 2002; Rodríguez-Jiménez and Murillo, 2011), Mexico (Carvallo, 2006; Blanco, 2008), Argentina (Cervini, 2009) and Chile (Mizala and Romaguera, 2003; Mizala, Romaguera, and Ostoic, 2005; Belleï, 2005), very few researchers have carried out longitudinal 'value added' (VA) studies by controlling for previous achievement. In this context, pioneer longitudinal VA studies have been published mainly in Argentina (Cervini, 2006) and Chile (Manzi, San Martín, and Van Bellegem, 2011; Carrasco and San Martín, 2011; San Martín and Carrasco, 2012; Thieme, Tortosa-Ausina, Prior, & Gempp, 2012; Thomas, Salim, Muñoz-Chereau, & Peng, 2012; Muñoz-Chereau, 2013). As Ferrer (2006) noted, the efforts to measure students' relative progress over time through value added measures represent isolated efforts made by some researchers in some specific Latin-American regions, but in most of the cases they have been discontinued before reaching visibility, let alone legitimacy of the national assessment system.

Thus, pioneering work has been carried out by Cervini (2006). He conducted a secondary analysis of students' transferring from primary to secondary schools using their 2003 examination data matched to their previous attainment in 2001 and survey results provided by the Directorate of Evaluation of Educational Quality in Buenos Aires, Argentina. The sample comprised 6,133 secondary students nested within 96 schools (for math outcomes) and 6,862 students nested within 97 schools (for language outcomes). Seven MLM models were employed, all identifying school and student levels. The overall goodness of fit of Model A (prior achievement only or VA) was moderate, explaining only 24.9% and 36.2% of the total variance in math and language, respectively; Model B (student level background factors only) explained 18.2% and 22.6% of the total variance; Model C improved the goodness of fit (prior achievement and student level background factors) to 34.2% and 43.2% of the total variance explained. The total school variance explained (or reduction in school level variation) also improved from 45% and 61.7% (Model A) to 50.5% and 71.3% (Model B); and to 75.7% and 76.0% (Model C) in math and language, respectively. In other words, the inclusion of prior achievement improved significantly the goodness of fit of the models. However, prior achievement (Model A) and individual characteristics of the students (Model B) together (Model C) provided the best model fit to estimate school effects in Argentina. The data analysis using further Models E, F, and G also showed that the incidence of student background factors aggregated at the school level (i.e. CVA) had a higher impact than the individual student background factors, which allowed him to conclude that the 'family heritage' of the individual student works through the peer grouping in schools because the 'culturally equal' tend to be segregated within the same institutions. The author argued that in this way contextual factors affected not only overall performance, but also progress and the idea that good measures of prior achievement can account for contextual effects is not applicable to Argentina (Cervini, 2006). According to this author, school 'composition' factors add highly significant power in explaining differences between schools and therefore should be considered in the analyses, a point followed up in more detail below.

Carrasco and San Martín (2011) also explored the use of VA models to assess market-driven mechanisms in secondary education operating in Chile, such as parental choice. In order to take account of possible exogeneity of prior attainment (Manzi, San Martín, and Van Bellegem, 2011), they focused only on Year 10 maths students that moved schools from 2004 to 2006, reducing the sample from near 200,000 to 79,073 students nested within 601 schools. Contrary to what might be expected, after carrying HLM VA analysis controlling for previous attainment (maths in Year 8), they reported no significant differences in schools' VA estimates between different parental school choice preferences for municipal/public, semi-private subsidised or private secondary schools. This indicates that parental pressure through school choice (demand side) in Chile has not had the effect on school effectiveness claimed by the voucher supporters. Moreover, finding no statistical VA differences between municipal/public and semi-private subsidised schools implies that school administrative dependence is independent from its performance. However, they also pointed out that when comparing Raw, CA and VA estimates, public/municipal schools appeared as proportionally more effective than when judged under CA and VA scores, whereas the opposite situation was observed in the subsidized schools. In other words, Raw scores (as an indicator of school performance routinely applied in Chile through SIMCE) have been misleading: they have not just overestimated subsidized schools' effectiveness, but also underestimated public schools effectiveness. Some public schools in Chile attain a low percentage of proficiency, but their students are making good progress, which is not given credit under an unadjusted raw score measure. Against the vision that state schools are inefficient, the results showed that public schools are typically helping their students make greater-than-average progress, and that the introduction of VA in the accountability system is particularly relevant for schools working in disadvantaged contexts.

Contextualized Value Added (CVA) Measures

In line with Cervini's findings (2006) among others, it has been stressed that CVA is a way of estimating the effect of school policies upon student achievement by explicitly controlling for prior achievement and compositional or contextual factors (Ballou, Sanders, & Wright, 2004). So the difference between VA and CVA is that the latter is an extension of VA that explicitly controls for intake differences between schools in pupils' academic and background characteristics, as well as context or compositional factors (Leckie &

Goldstein, 2011). Precisely because it takes into account contextual factors - and, critically, peer group effects that relate, for example, to the combined influence of SES over and above the individual effect - CVA is considered a methodologically more advanced or refined version of VA (Thomas & Mortimore, 1996; Rosenkvist, 2010; Timmermans et al., 2011). However, Harker & Tymms (2004) have noted that additional caution is required in the model specification and predictor reliability when constructing models that involve compositional effects. It is also claimed that CVA models are better for identifying those schools with a greater proportion of students from disadvantaged backgrounds that are able to boost student performance (OECD, 2008) and many studies have shown, the inclusion of individual student background/ socio-economic characteristics has helped fine-tune VA measures (Sammons, Thomas, Mortimore, Owen, & Pennell, 1993; Thomas & Mortimore, 1996; Sammons et al., 1997). Particularly for a system of external school accountability like the Chilean one, it may arguably be much fairer to employ CVA as the main indicator of school performance. For example, when schools that serve large concentrations of disadvantaged students are considered —like public schools— if they do not have sufficient compensatory resources to offset the educational challenges that such students pose, they may wrongly appear more ineffective despite using their insufficient resources more productively and efficiently than other schools (Ladd & Walsh, 2002). Interestingly, from 2006 CVA indicators -that take into account not only pupil prior attainment but also other pupil (gender, age, deprivation) and school context characteristics associated with performance differences outside the school control - were published for all schools in England as a central part of the school accountability system as well as for school improvement initiatives at the local and national levels (Ray, 2006; Evans, 2009). However, in 2010 CVA measures were subsequently removed from the English national school indicator system due to a government perception that these measures may in some cases lower expectations of school and student performance. Thus in practice, CVA approaches to measuring school effectiveness have not reached a consensus of support and further exploration comparing different approaches is required in Latin-America and different country contexts to inform policy development.

Contextualized Value Added (CVA) studies carried out in Latin-America and the Caribbean region

In Chile, Muñoz-Chereau (2013) carried out a multilevel study looking at the range and extent of school performance in secondary schools using CVA and Raw measures in language and math. The analyses involved a sample of 176,896 students (nested within 2,283 schools) that took the SIMCE tests in grades 10 and 8 (2006 and 2004) and their family questionnaires. The main findings from a 2-level analysis (school, student) reported large and statistically significant differences between the estimates of Chilean secondary schools' in different subject outcomes. The percentage of total variance in student's 'Raw' attainment attributable to differences between schools was sizable (36.9% in language; 47.3% in math) and interestingly considerably higher than similar primary school 'Raw' effects reported by Murillo and Roman (2011). However, after controlling for aspects arguably out of the school control, the equivalent CVA figures were dramatically reduced to 6.1% in language and 13.4% in math. Nevertheless, there remained a clear school effect demonstrated by the intra-school correlations.

These results are especially relevant because the overall goodness of fit of the CVA Model - which controlled for Prior attainment, Pupil characteristics/Background and School mean prior attainment - was fairly high, explaining 65.9% of the total and 94.2% of the school variance in language as well 69.1% of the total and 91.2% of the school variance in math. In other words, CVA Model was powerful enough to explain most but not all of the differences in language and maths achievement in Chilean secondary schools. Particularly, the precision gained by the CVA Model compared to a VA Model when measuring school performance was slightly better in both subjects, improving the goodness of fit from 65% to 66% in language, and from 65% to 69% in math, in terms of total variance explained. These results suggested that the inclusion of school context (mean prior attainment) helped to fine-tune the results when explaining differences between students in their relative progress and accounting for school level and student level variance.

Another CVA type study was carried out in Chile by San Martín and Carrasco (2012) using SIMCE data in Math and a more limited set of explanatory variables. They analysed a sample of 153,689 students nested in 3,203 public and mixed secondary schools (excluding schools in the private sector) that took SIMCE tests in 2005 and 2009 and reported estimates for each type of school. Using HLM, they developed a CVA model controlling for pretest (SIMCE score in 2005), school SES, school selectivity (based on student ability and the mean score from the school that each student went to in 2005). They compared the school CVA results with the model routinely used in Chile for judging school performance (based on SIMCE raw scores) and warned about the danger of using a biased approach such as the latter to classify schools, especially because more than the 60% of the analysed schools changed their effectiveness status from one approach to the other.

Finally, at the primary level Thieme et al. (2012) also employed a CVA approach. The analyses involved a sub-sample of 47,076 Year 8 (13 years-old) students (representing only 33% of the initial sample) nested within 948 Chilean primary schools (395 public, 460 mixed and 93 private) that took the 2009 SIMCE tests in language and math. This data was matched to students' prior attainment in the same test four years before (in 2005) and their family questionnaires. The analysis compared 4 models: Raw (empty model), CA (controlling for SES index - created by the authors including parents' education and family income- at the student level and aggregated at the school level), VA (prior attainment only: SIMCE 2005 in language and math), and CVA (prior attainment SIMCE 2005 in language and math; student background characteristics (student SES index) and school context (SES index aggregated at the school level)). They reported an Intra School Correlations (ICC) ranging from 34% (Raw) to 12% (CVA), fairly similar to equivalent ICCs reported by Muñoz-Chereau (2013) at the secondary level. In line with previous research they also found a strong positive correlation between the school residuals obtained from Raw and CA models (0.78), but a much lower correlation between Raw and CVA models (0.44). However, the strong data reduction (67% of the initial sample) and omission of relevant information such as the percentage of the total/school variance explained by the models and results for each subject analysed (as well as the fact that they did not refer to similar work done by other researchers in Chile) imposes limitations to their study.

To summarise, Raw and CA studies conducted in Latin-America have highlighted bigger primary school effects than those typically identified in developed countries which have fluctuated between 6% and 20% (Smith and Tomlinson, 1989; Fitz-Gibbon, 1991; Tymms, 1993; Bosker and Witziers, 1996; Sammons, Thomas, and Mortimore, 1997). Multilevel studies carried out in the region have also identified country effects and stressed the need to consider regional and class differences when evaluating national policy and practise, even though few studies exist internationally on this topic. Considering that 'regional context (such as socioeconomic or geographical factors) or education policy (such as the extent of selection or private schooling) may limit the possibilities of a school being more or less effective, as well as enhance or inhibit the overall differences between schools' (Op.cit, p.314) regional as well as class effects (Goldstein, 1997; Martínez, 2012) are likely to play a very significant role in Latin-America and the Caribbean and need to be taken into account when looking at school effects. Focusing on the few isolated Latin-American VA and CVA studies conducted up to now, Argentina and Chile stand out highlighting compositional or contextual factors (probably due to segregation of their educational systems), justifying the need of using these type of models for fairer and more accurate comparisons of school performance.

However, the lack of longitudinal data including previous achievement in the Latin-American studies as well as the samples mostly being at the primary school level impose limitations regarding the conclusions that can be drawn in relation to the effectiveness of secondary schools and the extent of the regional and class differences. Hence this study seeks to clarify previous findings by analysing student attainment data and school effects using four different approaches (Raw, CA, VA and CVA) but also by specifying different levels (2, 3 and 4) in the analyses according to municipality, school, and class identifiers. The key issues we are seeking to address are the relative importance of each level as a source of achievement variation as well as the need to include different predictors in the analysis so as to provide new evidence for developments in education policy and practice. We address the following specific research questions: (1) Does the estimation of school effects in students' language and mathematics performance change when four typical approaches to measuring school effectiveness (Raw, CA, VA and CVA) are used?; (2) Does the inclusion of different 'clustering' levels (2, 3 or 4) change the estimation of school effects in students' language and mathematics performance?; and (3) Do these results mirror or modify the conclusions obtained by similar studies conducted in developed countries?

Methodology

Data and Sample

The longitudinal sample comprised the 2006 cohort of 177,461 grade 10 students, nested within 7,146 classrooms, 2,283 secondary schools and 313 municipalities that took the National System of Measurement of Educational Quality tests (SIMCE) in two subject outcomes (language [Spanish] and maths). This data was also matched to students 2004 SIMCE prior attainment in grade 8 (end of primary schooling) and pupil/family characteristics comprising the variables: gender, mothers education, number of books in the home and number of people in the home. See Tables 2 & 3 and Muñoz-Chereau (2013) for further details of the sample and variables employed.

Modeling Approach

In order to explore the extent to which pupil intake and background features of Chilean secondary schools influence students' performance in mathematics and language, four types of models identified in the literature review (Raw, CAM, VA and CVA) are employed to explore and compare estimates of secondary school effectiveness in Chile. MLwin software is employed for all analyses (Rasbash et al, 2005). For each student outcome measure (mathematics and language) the four models are applied separately and adjust for the following explanatory variables in the fixed part of the model:

Raw model: no explanatory variables

CA model: Pupil characteristics/background (Gender, Mother's education, Number of Books in the home and Number of people in the home)

VA model: Pupil prior achievement (Language Year 8, Math Year 8)

CVA model: Pupil prior achievement (Language Year 8, Math Year 8) plus Pupil characteristics/background (Gender, Mother's education, Number of books in the home and Number of people in the home) plus School context (Language Year 8 school mean or Math Year 8 school mean, respectively)

The individual and family characteristics employed in the CA and CVA model were identified in previous research as the key socio-economic variables for the purpose of these analyses (Muñoz-Chereau, 2013). Moreover, each of the four model analyses (Raw, CA, VA and CVA) are repeated using three different 'clustering' specifications in the random part of the ML models: 2-level (students within schools), 3-level (students within classrooms within schools) and 4-level (students within classrooms within schools within municipalities). This MLM technique decomposes the variation in student language and mathematics outcome scores, unexplained by the explanatory variables in the fixed part of the models, into variance components that can be attributed to each of the different levels specified in the random part of the models.

The findings are reported in tables which provide: (1) estimates of the fixed part predictors and random part variances obtained for each of the four models (Raw, CAM, VA and CVA) employed. However, due to the similarity of fixed part estimates when different clustering levels are specified only the 4-level model results are reported (see Tables 2 and 3); (2) for each model employed the percentage of variance explained at each level and the intra-class correlations (ie that reflect the percentage of variance that is attributable to each level specified in the model) (see Tables 4 and 5) and (3) categorisation and comparison of school residuals from 2 and 4 level models.

Findings

Research Question 1: Does the estimation of school effects in students' language and mathematics performance change when four typical approaches to measuring school effectiveness (Raw, CA, VA and CVA) are used?

As would be expected from previous research outlined in the Literature review, the results from all model analyses in Tables 2 and 3 show that across both language and maths outcomes the impact on student Year 10 attainment of all explanatory variables included in CA, VA and CVA models were statistically significant (at 0.05 level). Not surprisingly the prior attainment scores relevant to the outcome analysed (either language or math in Year 8) have the greatest impact. In terms of gender, the average Year 10 score difference between boys and girls was found to be larger in maths than language using the CVA model; 4.7 points (language) and -7.2 points (maths). Moreover, the equivalent figures for CA model were 6.1 points (language) and -13.3 (math), indicating that when absolute attainment is considered rather than progress, gender differences in favour of girls (language), and boys (maths), are more extreme in both subjects. Mothers' education, number of books in the home and number of people in home are measures that reflect students' family socio-economic status. Again not surprisingly given previous research, language outcomes are more influenced by family characteristics than maths outcomes. Specifically, students' language attainment and progress (CA and CVA models) is almost always higher in relation to greater "mothers' education", more "books in household" and fewer "people in the household". Whereas students' maths attainment and progress (CA and CVA models) is only clearly higher in relation to more "books in the household". The equivalent findings for the two other family factors "mothers' education" and "number of people in the household" are not conclusive for students maths outcomes, given most estimates for sub-categories were not statistically significant and quite surprisingly three statistically significant estimates were found to be negative, thereby indicating the possibility of confounding between socio-economic family factors. Finally, in terms of school context (school mean prior attainment in Year 8

Language or maths), this is found to have a positive association with students' absolute attainment and value added performance, comparable to the influence of individual level prior attainment in the same subject. This indicates that the extremes of greatest and least progress for individual students are influenced by their school peer group and typically enhanced by relatively higher mean prior attainment of all students in the school year group.

Insert (Table 2) and (Table 3) here

Having accounted for the explanatory variables in the fixed part of the models, it is clear that there is a considerable reduction of the total variance in student outcomes that is attributable to schools from Model 1 (Raw) to Model 4 (CVA). For example using the 4-level model the intra-school correlation drops from 37.3% to 9.1% in Math and 14.7% to 4.0% in Language (see Tables 4-5), after accounting for pupil prior achievement, pupil characteristics/background and school context. Nevertheless, for the CVA model there is still significant variation between Chilean secondary schools remaining after taking account of intake factors and these differences are much more pronounced for math than language.

With regard to the best model for estimating school effects, the findings indicate that for both language and maths outcomes CVA models provide the best goodness of fit, irrespective of the number of clustering levels employed. For example, using the 4-level models, the percentage of total variance explained for VA and CVA improves from 61% to 68% for maths and from 58% to 62% for language. These findings are in line with the few existing previous studies conducted in the region that have pointed out that CVA provided by far the best model to estimate school effects (Cervini, 2006; Thomas, Salim, Muñoz-Chereau, and Peng, 2012; Muñoz-Chereau, 2013). Nevertheless, it would be valuable to conduct further research using similar VA and CVA models and data from further student cohorts to build a stronger evidence base indicating the impact of school context factors on student attainment and progress in Chilean secondary schools. Overall, for the purpose of holding schools accountable for their students' results, adjusting school performance measures for student characteristics/background and school context as well as student prior achievement, is recommended by both this study and previous research (Leckie and Goldstein, 2011; Timmermans, op. cit.) and seems also appropriate for Chile.

Research Question 2: Does the inclusion of different 'clustering' levels (2, 3 or 4) change the estimation of school effects in students' language and mathematics performance?

The findings clearly show that the estimation of the relative importance of school effects in Chile are determined by the levels specified in the model, in particular by the inclusion of lower (class) and upper levels (Municipality). For example, from the Language CVA model the variance attributable to schools ranged from 6.1% (2-level model) to 4.3% (4-level model) and for maths the equivalent figures are 13.2% (2-level model) to 9.1% (4-level model). Indeed irrespective of the type of model employed (Raw, CA, VA or CVA) the school effect is inflated in the absence of class and municipality levels. For the maths outcome 2-level models, the Intra-school correlations ranged from 47.3% (Raw) to 13.2% (CVA) and the equivalent figures are reduced for the 4-level models, 37.3% (Raw) to 9.1% (CVA). For language outcome 2-level models, the Intra-school correlations ranged from 36.9% (Raw) to 6.1% (CVA) and the equivalent figures are similarly reduced, but to a great extent, for the 4-level models, 14.7% (Raw) to 4% (CVA).

Insert (Table 4) and (Table 5) here

With regard to the estimated size of Municipality and class effects, in most cases these are smaller than the school effects, but nevertheless of substantive importance. For example, using a 4-level model the variance attributable to Municipality ranged from 6% (Raw) to 0.3%-0.5% (CVA) for maths and Language outcomes respectively. Focusing on the variance attributable to differences between classes (within schools), again using a 4 level model, the variance attributable to classrooms ranged from 9-13% (Raw) to 5-10% (CVA) for maths and Language outcomes. It is also clear that for Raw, CA and VA models, the variance attributable to classroom differences appears consistently lower than the school variance, irrespective of the clustering levels. However, importantly for the CVA (3 and 4 level) models, the variance attributable to classroom appears slightly higher than the school variance. Overall, these results indicate that even though at least twice as much of the total variance is attributable to the class level than the Municipality level in both subjects, any fair effort to assess the quality of schools in Chile should not neglect the Municipality level because the regional context (such as socioeconomic or geographical factors) is also significantly affecting the possibilities of a school of being more or less effective.

This argument is supported and emphasised to an even greater extent when the school residuals from CVA models employing different clustering levels are categorised and compared (see Tables 6 and 7). The results show how the apparent precision assumed by a 2 level model in drawing statistically significant conclusions about the effectiveness of some Chilean secondary schools could be misleading, when in contrast, by using a 4 level model no statistically significant school residuals may be indicated. For maths and language outcomes respectively, there are 423 (19%) and 274 (12%) schools that would be classified as performing statistically significantly lower than expected using a 2 Level CVA Model but performing as expected using a 4 Level model. At the same time, there are 387 (17%) and 263 (12%) schools that would appear as performing statistically significantly higher than expected ('adding value') using a 2 Level CVA model, but as expected under a 4 Level model. Overall, for language and maths outcomes respectively, 564 (25%) and 1029 (45%) schools would be classified differently when using a 2 Level CVA Model in comparison to a 4 Level Model.

Crucially, this suggests that by employing a simpler 2-level CVA model, one in four (25%) schools using language outcomes and almost one in two (45%) schools using math outcomes may be misclassified as having a value added result either lower than expected, as expected or higher than expected. Furthermore, using a more sophisticated and rigorous 4-level model is likely to indicate far fewer schools are performing differently from what might be expected in comparison to the 2-level model. For those schools potentially misclassified - as well as for the school system as a whole - this of course could have serious high stakes consequences.

Insert (Table 6) and (Table 7) here

Importantly, these findings are in line with and also extend previous studies carried out in the region (Cervini, 2006; Thomas, Salim, Muñoz-Chereau, and Peng, 2012) and strongly support the need to recognise all relevant 'clustering' factors in the multilevel analysis of school performance. This approach is required not only to address violations in the 'independent observations' assumption within MLM clusters but also to evaluate and recognise the important potential impacts of regional/Municipal and classroom factors on student attainment and progress.

Research Question 3: Do the results for Research Questions 1 and 2 mirror or modify the conclusions obtained by similar studies conducted in other countries?

The results obtained by this study regarding Research Question 1 are in line with previous research conducted in Chile and elsewhere, especially when pointing out that CVA models typically provide the best fit of the data (in terms of total variance explained) and that differences between schools remain, although to a lesser extent, after student background characteristics, school contextual factors and prior achievement have been taken into account (Thomas, 2001; Opdenakker et al., 2000, 2002, 2006; Muijs and Reynolds, 2003; Webster and Fisher, 2000, Cervini, 2009).

Moreover, when the 2-level model (schools, students) is employed to compare Chilean CVA study results to those found in other countries, the findings are fairly similar in terms of the range and extent of secondary school effects for both subjects (intra-school correlation: 6.1% [Language] and 13.2% [maths] see Tables 4 and 5). In England, for example, the percentages of remaining variation in students GCSE outcomes attributable to differences between schools for English and mathematics when using CVA models have been reported as 9% and 12%, respectively (Thomas & Mortimore, 1996; Thomas, 2001). These findings from England were recently updated by Thomas & Peng (2013) using the same longitudinal datasets but this time in terms of GCSE total score for three separate years (1993, 1999, 2006) over a fourteen year period; equivalent intra-school correlations for CVA models were reported and ranged from

7% to 9% (raw models: 14%-21%; VA models: 9%-13%). Similarly, in a different country context – China - equivalent intra-school correlation results for CVA models were also reported by the same authors from a Chinese longitudinal dataset of senior secondary schools for three consecutive years (2009, 2010, 2011). In this case the student outcome measure was total entrance examination to higher education score and the sample was two local authorities, one in the poorer west and one in mid/east of China. The findings for CVA models ranged from 4% to 14% (raw models: 22%-27%; VA models: 15%-22%) in the poorer western LEA with 30 schools. The equivalent figures for the more affluent mid/east LEA were in most cases lower 4% to 7% (raw models: 22%-24%; VA models: 8%-9%) (Thomas & Peng, 2013) and broadly in line with the raw between school variance in PISA results reported for the very affluent Shanghai region (OECD, 2013 – see Figure II.2.7).

Thus overall, when the CVA school effect estimated in this study is compared with the same effect reported in other countries - particularly England (Thomas and Mortimore, 1996), China (Thomas and Peng, 2013) and the Netherlands (Timmermans, op. cit.) - it can be concluded that Chilean school effects are fairly similar to the ones reported in these countries. However, the time period between prior and outcome attainment examined in Chile is typically shorter than in other country contexts (eg two years in Chile vs. five years in England), suggesting that Chilean estimates may be conservative and at the lower end. As mentioned previously, further research using more than one student cohort would be beneficial to replicate these findings over time as well as to consider a separate important issue not addressed in the current study - the apparent lack of stability over time in school effects (Thomas et al., 2007).

With regard to Research Question 2 the results obtained by this study are also in line with previous research conducted in Chile and elsewhere, but crucially go further than earlier studies in demonstrating the key importance of taking into account both class and municipality clustering factors in the analysis, as well as the school level. Unfortunately, few studies exist with sufficiently detailed data to conduct such analyses but we discuss the most relevant of these below in relation to our findings. Blanco (2008) conducted a multilevel study analysing a representative sample of 51,053 Mexican Year 6 students' achievements in the 2003 national tests in Language and Math, nested within 2,752 primary schools. Although a 3 level nested model structure was used (pupils within schools within states), the number of upper level units was not reported. Aside from this limitation, raw scores differences between schools and states were reported to account for 28% and 6.5% of the total variance in pupils' language scores. The equivalent figures for Math were 25% and 3.5%, slightly lower. Moreover, Cervini (2009) analysing Argentinian 6th grade students mathematics attainment from the 2000 national school census has indicated that the estimation of the relative importance of primary school effects in Argentina are strongly affected by the levels specified in the model, in particular the lower (class) as well as the upper (municipality and state). Using a model that controls for prior attainment and student background factors he concludes that the class level has the largest impact on the estimation of the school effect. By omitting it, there is an overestimation of school effects and the associated intra-school correlations (the percentage of variance attributable to the class, school, municipality and state effects is 11.5%, 12.0%, 2.5% and 2.9%, respectively). The author suggested that this might be the case due to the existence of strong contextual effects, although this study did not use prior achievement measured in a standardized test, but prior teacher assessed grades and repetition of school year as predictors, imposing limitations on the findings. Interestingly, the Cervini findings from a Latin American context are slightly larger but broadly similar to the equivalent estimates reported in the current Chilean study, even though we benefited from more reliable standardised prior attainment measures. With regard to similar research that has investigated class, school and regional effects in developed country contexts such as the UK, again little comprehensive research exists. Of the evidence available, typically class level effects are often argued to be larger than school effects (Rowe, 2003), and school effects are larger than regional or LEA effects (Thomas, 2001). The CVA results of this study are generally in line with previous research and with the statement that "there is more variation among classrooms within schools than there is among schools [...] a significant shortcoming of [international] studies is the lack of data at the classroom level" (Willms, 2006, p. 55). However, it is pertinent to emphasise that in the current study CVA class effects are actually found to be only slightly higher than school effects, and indeed are smaller than school effects for Raw, CA and VA models. This supports similar findings of Thomas (2001) using Scottish data and one implication is that in some educations systems, due to the way schooling is organised, segregation may be much more apparent between schools, than between classes within schools. Overall, in spite of the relative size of class, school and municipal effect, the key finding remains - that the extent and significance of school residuals changes depending on the clustering levels controlled for in the analysis.

Discussion

By situating these findings in the broader knowledge-base of Educational Effectiveness Research via comparison with similar studies carried out in the Latin-American region as well as in other countries, several key points can be highlighted. First, the evidence from this study supports and extends previous similar research (eg Muñoz-Chereau, 2013; Timmermans & Thomas 2014; Ballou, Sanders & Wright, 2004; Thomas 2001) and finds that the estimation of the relative importance of school effects are affected by both the modelling approach taken to measure school effectiveness (Raw, CA, VA or CVA) and the outcome analysed (language or math). Overall, the CVA/VA approach provides fairly similar estimates of secondary school effects in Chile and other countries (UK, China). In contrast and not surprisingly raw school performance tends to vary to a somewhat greater extent in poorer and unequal societies such as Chile and China (especially poorer western China) than in UK. Although it should be noted that in previous UK studies raw variation in student performance is typically under-estimated as private schools often do not provide the required data and are therefore excluded from analyses. The findings regarding Chilean raw school performance in this study are also reflected by other evidence that secondary schools in Chile appear less effective in some aspects. For example, in terms of average scores in international comparative surveys, as well as in the extent of school variation in raw attainment scores (ie intra-school correlation for raw scores) and in the size of the equity gap (OECD, 2010).

Secondly, school effectiveness studies that do not take account of clustering levels below and above the school, may be mis-specified and misleading, tending to overestimate school effects. This study was able to identify new evidence of substantial and statistically significant Municipality, school and classroom effects within Chilean secondary school system and the findings typically support the limited previous research on this topic (eg Cervini, 2009). In other words, all of these clustering levels matter when explaining pupil attainment and progress in secondary schools in Chile and this issue is also stressed by the increasing shift in focus and terminology used in the literature from “school effectiveness” to “educational effectiveness” (Chapman et al, 2012). Thus a key conclusion from this study is that a 4 level model (Municipality, schools, classes and students) provides more sophistication in the analysis and better precision and explanatory power in estimating school effects than the more common two level model that focuses only on students within schools. The argument for this is provided by the evidence that there is significant variance at the Municipality and Classroom levels. Findings from different education systems also suggests that the macro-level, that is the national or regional policy level, plays a salient role in what happens in the classroom, a conclusion already outlined with regard to other countries when detecting variability across regions in the percentage of variance explained by different factors (Thomas, 2001) and in raw performance variance attributed to regional differences (OECD, 2013 – see Figure II.2.a). Therefore in the presence of a high stakes accountability policy such as the one under implementation in Chile, this study advocates the use of a 4-level model when estimating school effects, because there are regional/Municipality and class differences that need to be taken into account in order to develop fairer measures to compare schools.

Thirdly, in line with Fertig and Schmidt (2002), this study supports the view that EER in developing countries such as Chile needs to move towards a more contextual model which takes account not just of individual student characteristics or process factors, but more importantly, the socio-economic, political and cultural dimensions of the school and the broader regional and national context. In order to be able to differentiate whether children learn more or less because of the policy (arguably out of the school control) or practice (arguably under the school control,) as argued by Sammons (2007) among others, the approach of mainly focusing on practices at the school level may be misguided when trying to disentangle the complexities in different national and regional/Municipality contexts such as the one described in this study. Given that the trend of evaluating school effects based on detailed analyses of student assessment data is increasingly being used as a key strand of school inspection and self evaluation systems across the world, a relevant policy and practical implication supported from these findings is that different types of school performance measures may serve different policy purposes. If the focus is to evaluate schools on their improvement and inform school improvement initiatives, a need to adjust for factors both jointly and separately, such as student background, context, input, process (arguably either within or outside the control of school) – may be required in order to investigate school performance in detail and provide teachers with a variety of relevant feedback to improve their practice. However, if the emphasis is on school choice, the most appropriate approach may only need to adjust for prior attainment, but to provide separate measures for different student groups (i.e girls/boys, low/high achievers, etc) - as is currently reported in English school performance tables (DfE, 2014). But for a policy oriented to support accountability mechanisms with high stakes consequences for schools, a contextualised value added (CVA) model that adjusts for student background factors and compositional or context effects, outside of the control of school, as well as prior attainment may be most appropriate. This approach, in spite of debatable limitations (Timmermans and Thomas, 2014; Harker & Tymms, 2004), has the potential to inform and enhance educational policy, programme development and practice but in a fairer way that acknowledges the extremely advantaged or disadvantaged context of some schools, especially those in developing countries. This approach may also be viewed as more conservative because estimates of CVA school effects are typically less variable than other performance measures. The high stakes accountability system currently being implemented in Chile will have very serious consequences for some schools (eg resulting in school closures) so taking a cautious more conservative approach seems especially fit for the purpose in this context.

Table 1: Country classification across SERCE evaluated subjects and Primary grades 3 and 6

	Countries that exhibit mean scores significantly higher than the regional average	**Countries matching the regional average**	**Countries that exhibit mean scores significantly lower than the regional average**
Math 3	Chile, Costa Rica, Cuba, Mexico and Uruguay	Argentina, Brazil and Colombia	Dominic Republic Ecuador, Guatemala, Nicaragua, Panama, Paraguay, Peru and El Salvador.
Math 6	Argentina, Chile, Costa Rica, Cuba, Mexico and Uruguay	Brazil, Colombia and Peru	Dominic Republic Ecuador, Guatemala, Nicaragua, Panama, Paraguay and El Salvador.
Reading 3	Argentina, Chile, Colombia, Costa Rica, Cuba, Mexico and Uruguay	Brazil and El Salvador	Dominic Republic Ecuador, Guatemala, Nicaragua, Panama, Paraguay and Peru
Reading 6	Brazil, Chile, Colombia, Costa Rica, Cuba, Mexico and Uruguay	Argentina	Dominic Republic Ecuador, Guatemala, Nicaragua, Panama, Paraguay, Peru and El Salvador.

Table 2: 4-level Model: Language Year 10

4 levels	**Model 1 Raw**	**Model 2 CA**	**Model 3 VA**	**Model 4 CVA**
Variables	**Estimate [SE]**	**Estimate [SE]**	**Estimate[SE]**	**Estimate [SE]**
Fixed part (coefficients)				
Cons (Intercept)	259.885*[1.09]	243.808* [1.01]	258.544* [0.40]	117.32* [1.99]
Pupil prior achievement				
Language Year 8	NA	NA	0.52* [0.00]	0.50* [0.00]
Mathematics Year 8	NA	NA	0.22* [0.00]	0.23* [0.00]
Pupil characteristics/background				
Gender (girls vs boys)	NA	6.1* [0.23]	NA	4.7* [0.18]
Mothers' education				
Secondary incomplete (vs Primary)	NA	1.7* [0.40]	NA	0.27* [0.00]
Secondary complete (vs Primary)	NA	5.5* [0.36]	NA	1.59* [0.28]
Vocational/Secondary incomplete (vs Primary)	NA	7.25*[0.96]	NA	2.18*[0.73]
Vocational/Secondary complete (vs Primary)	NA	7.26*[0.57]	NA	2.29*[0.43]
University graduate incomplete (vs Primary)	NA	11.97*[0.99]	NA	3.45*[0.76]
University graduate complete (vs Primary)	NA	12.89* [0.59]	NA	4.38*[0.45]
University postgraduate incomplete (vs Primary)	NA	5.32*[0.35]	NA	1.46*[0.27]
University postgraduate complete (vs Primary)	NA	-1.33*[0.37]	NA	-0.14[0.28]
Books in the house				
11-50 (vs 0-10 books)	NA	6.78* [0.26]	NA	1.65* [0.20]
51-100 (vs 0-10 books)	NA	11.76* [0.33]	NA	3.12* [0.26]
> 100 (vs 0-10 books)	NA	17.02* [0.38]	NA	4.59* [0.29]
People in the house				
4-6 (vs 2-3 people)	NA	-2.81* [0.30]	NA	-0.96* [0.27]
7 or more (vs 2-3 people)	NA	-4.87* [0.40]	NA	-1.97* [0.30]
School context				
Language Year 8 mean	NA	NA	NA	0.30*[0.00]
Random part (Variances)				
Between municipalities:				
Cons (Intercept)	180.335[27.15]	86.334[17.61]	12.356[3.10]	3.386[1.14]
Between schools:				
Cons (Intercept)	400.624[18.97]	578.972[22.16]	122.098[5.19]	41.858[2.53]
Between classrooms				
Cons (Intercept)	349.749[8.86]	187.743[5.36]	53.565[1.88]	51.69[1.88]
Between pupils:				
Cons (Intercept)				

	1786.949[6.13]	1604.609[5.75]	952.269[3.27]	942.303[3.38]
Total:	2727.657	2457.658	1140.288	1039.237

Note: *= statistically significant at 0.05 level

Table 3: 4-level Model Mathematics Year 10

4 levels	**Model 1 Raw**	**Model 2 CA**	**Model 3 VA**	**Model 4 CVA**
Variables	**Estimate [SE]**	**Estimate[SE]**	**Estimate[SE]**	**Estimate [SE]**
Fixed part (coefficients)				
Cons (Intercept)	250.41*[1.60]	251.574*[1.55]	256.75* [0.70]	110.818* [2.86]
Pupil prior achievement				
Language Year 8	NA	NA	0.18* [0.00]	0.19* [0.00]
Mathematics Year 8	NA	NA	0.68* [0.00]	0.65* [0.00]
Pupil characteristics/background				
Gender (girls vs boys)	NA	-13.31* [0.26]	NA	-7.16* [0.26]
Mothers' education				
Secondary incomplete (vs Primary)	NA	-0.61 [0.46]	NA	-0.07 [0.34]
Secondary complete (vs Primary)	NA	-0.15 [0.42]	NA	-0.22 [0.31]
Vocational/Secondary incomplete (vs Primary)	NA	-2.16 [1.12]	NA	-1.68*[0.84]
Vocational/Secondary complete (vs Primary)	NA	-1.16 [0.65]	NA	-0.79 [0.48]
University graduate incomplete (vs Primary)	NA	0.82 [1.14]	NA	0.83 [0.85]
University graduate complete (vs Primary)	NA	-1.54*[0.66]	NA	0.23[0.49]
University postgraduate incomplete (vs Primary)	NA	-0.35 [0.39]	NA	-0.34 [0.29]
University postgraduate complete (vs Primary)	NA	-0.56 [0.43]	NA	-0.37 [0.32]
Books in the house				
11-50 (vs 0-10 books)	NA	8.00* [0.29]	NA	1.78 *[0.22]
51-100 (vs 0-10 books)	NA	13.56* [0.37]	NA	3.04* [0.28]
> 100 (vs 0-10 books)	NA	19.59* [0.42]	NA	4.25* [0.32]
People in the house				
4-6 (vs 2-3 people)	NA	-0.46 [0.34]	NA	0.58* [0.26]
7 or more (vs 2-3 people)	NA	-1.75*[0.45]	NA	0.26 [0.34]
School context				
Mathematics Year 8 mean	NA	NA	NA	0.57* [0.01]
Random part (Variances)				
Between municipalities:				
Cons (Intercept)	291.425[54.82]	230.410[45.19]	40.966[9.74]	7.350[2.75]
Between schools:				
Cons (Intercept)	1659.657[59.58]	1468.86[53.45]	394.034[15.21]	130.098[6.73]
Between classrooms				
Cons (Intercept)	404.916[10.12]	390.890[9.95]	144.965[3.97]	145.023 [4.00]
Between pupils:				

Cons (Intercept) Total:	2089.540[7.17] 4445.538	2025.059[7.27] 4115.219	1164.922[3.99] 1744.887	1144.404[4.11] 426.875

Note: *= statistically significant at 0.05 level

Table 4: Percentage of Variance explained and Intra-class correlations: Language Year 10

2-level Model Language	**Model 1 Raw**	**Model 2 CA**	**Model 3 VA**	**Model 4 CVA**
% variance explained School Pupil Total	 NA NA NA	 32% 2.9% 13.6%	 86% 44.7% 60%	 94% 45% 63.3%
% variance attributable to School Pupil	 36.9% 63.1%	 29.1% 70.1%	 13% 87%	 6.1% 94%
3-level Model Language				
% variance explained School Classroom Pupil Total	 NA NA NA NA	 30.3% 11.6% 2.1% 12.5%	 86% 75% 42% 59.2%	 95.3% 75.6% 42.5% 63.1%
% variance attributable to School Classroom Pupil	 34.3% 7.5% 58.2%	 27.3% 7.6% 66.1%	 11.8% 4.7% 83.4%	 4.3% 5% 91%
4-level Model Language				
% variance explained Municipalities Schools Classrooms Pupils Total	 NA NA NA NA NA	 52.1% 44.5% 48% 10.2% 9.9%	 93.1% 69.5% 85.1% 47% 58.2%	 98.1%% 89.5% 85.6% 47.2% 62%
% variance attributable to Municipality School Classroom Pupil	 6.6% 14.7% 13.2% 65.5%	 3.5% 23.5% 7.6% 65.3%	 1.1% 10.8% 4.7% 83.5%	 0.3% 4% 5% 91%

Table 5: Percentage of Variance explained and Intra-class correlations: Mathematics Year 10

2-level Model Math	**Model 1** **Raw**	**Model 2** **CA**	**Model 3** **VA**	**Model 4** **CVA**
% variance explained				
School	NA	21.2%	78.6%	91.1%
Pupil	NA	3.5%	46.8%	47.6%
Total	NA	11.9%	61.8%	68.2%
% variance attributable to				
School	47.3%	42.3%	26.5%	13.2%
Pupil	52.6%	57.6%	73.5%	86.8%
3-level Model Math				
% variance explained				
School	NA	17.8%	77.8%	93.1%
Classroom	NA	5.8%	64.2%	64.2%
Pupil	NA	3.2%	44.2%	45.2%
Total	NA	10%	61%%	68.1%
% variance attributable to				
School	44.3%	40.4%	25%	9.6%
Classroom	9%	9.4%	8.3%	10.2%
Pupil	46.7%	50%	67%	80.2%
4-level Model Math				
% variance explained				
Municipality	NA	21%	86%	97.5%
School	NA	11.5%	76%	92.2%
Classroom	NA	3.5%	64%	64.2%
Pupil	NA	3.1%	44.2%	45.2%
Total	NA	7.4%	61%	68%
% variance attributable to				
Municipality	6.5%	5.6%	2.3%	0.5%
School	37.3%	35.7%	22.6%	9.1%
Classroom	9.1%	9.5%	8.3%	10.2%
Pupil	47%	49.2%	66.8%	80.2%

Table 6: CVA School Residuals categorised and compared for 2-level and 4-level models -Language Year 10

		CVA Two_Level Model			Total
		School residual statistically significant lower than expected	School residual as expected	School residual statistically significant higher than expected	
CVA Four_Level Model	School residual statistically significant lower than expected	4	12	0	16
	School residual as expected	274	1714	263	2251
	School residual statistically significant higher than expected	1	14	1	16
Total		279	1740	264	2283

Table note: Confidence interval used to categorise school residuals as higher, lower or as expected = +/- 1.96 s.e.

Table 7: CVA School Residuals categorised and compared for 2-level and 4-level models - Mathematics Year 10

		CVA Two_Level Model			Total
		School residual statistically significant lower than expected	School residual as expected	School residual statistically significant higher than expected	
CVA Four_Level Model	School residual statistically significant lower than expected	30	87	29	146
	School residual as expected	423	1194	387	2004
	School residual statistically significant higher than expected	22	81	30	133
Total		475	1362	446	2283

Table note: Confidence interval used to categorise school residuals as higher, lower or as expected = +/- 1.96 s.e.

References

Aitkin, M and Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society.* Series A (General), Vol. 149, No. 1, pp. 1-43.

Ballou, D., Sanders, W. and Wright, P. (2004) Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Educational and Behavioural Statistics*. Spring, Vol. 29, No. 1, p. 29: 37.

Belleï, C. (2005), *The private-public School controversy: the case of Chile*. Cambridge, Massachusetts, John F. Kennedy School of Government, Harvard University. at URL: http://www.periglobal.org/sites/periglobal.org/files/BELLEI%20-PEPG-05-13%20Private%20Public%20in%20Chile.pdf (viewed 7/03/13).

Belleï (2013) 'Situación Educativa de América Latina y el Caribe: Hacia la educación de calidad para todos al 2015'. at URL: http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Santiago/pdf/situacion-educativa.pdf (viewed 7/03/13).

Bosker, R. & Scheerens, J. (1989) Issues in the Interpretation of the Results of School Effectiveness Research. *International Journal of Educational Research* 13: p.741-751.

Bosker, R. and Witziers, B. (1996) *The magnitude of the school effect. Or: does it really matter which school a student attends?* Paper presented at the Annual Meeting of the American Educational Research, New York, NY.

Blanco, E. (2008) Factores escolares asociados al aprendizaje en la educación primaria mexicana: un análisis multinivel. REICE-*Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación.* Vol. 6, N.1.

Braun, H., Chudowsky, N. and Koenig, J. (2010). *Getting Value Out of Value-Added.* Report of a Workshop. Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Accountability. Center for Education Division of Behavioural and Social Sciences and Education. NATIONAL RESEARCH COUNCIL and NATIONAL ACADEMY OF EDUCATION. at URL: http://www.nap.edu/catalog/12820.html (viewed 24/01/13).

Carvallo, M. (2006). Factores que afectan el desempeño de los alumnos mexicanos en edad de educación secundaria: un estudio dentro de la corriente de eficacia escolar. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, REICE*, 4 (3), 30-53. at URL: http://www.rinace.net/arts/vol4num3/art3.htm. (viewed 7/03/13).

Carrasco, A. and San Martín, E. (2011). *Are quasi-markets in Education Meeting their Policy Purposes in Chile? Re-examining Empirical Hypothesis from Value-Added Models*. at URL: mideuc.cl/wp-content/uploads/2011/09/it1104.pdf (viewed 12/10/12).

Casas, A., Gamboa, L., & Piñeros, J. (2002). El efecto escuela en Colombia 1999 - 2000. *BORRADORES DE INVESTIGACION* 002499, UNIVERSIDAD DEL ROSARIO.

Cervini, R. (2006). Progreso de Aprendizaje en la educación secundaria básica de Argentina: Un análisis multinivel de valor agregado. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*. Vol.4, N.3. at URL: http://www.rinace.net/arts/vol4num3/art4_htm.htm (viewed 12/10/12).

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

Chapman, P. Armstrong, A. Harris, D. Muijs, D. Reynolds and P. Sammons (Eds.) (2012) *School Effectiveness and School Improvement Research, Policy and Practice. Challenging the Orthodoxy?*. London, Routledge pp. 125-145.

Creemers, B. (1996) The goals of school effectiveness and school improvement. In D. Reynolds, R. Bollen, B. Creemers, D. Hopkins, L. Stoll and N. Lagerweij (Eds). *Making good schools: Linking School Effectiveness and School Improvement*. London: Routledge pp.36-58.

Creemers, B. (1997). *Effective schools and effective teachers: An international perspective.* Centre for Research in Elementary and Primary Education, University of Warwick.

Evans, H. (2009) *Value-Added in English Schools*. Dept for Children, Schools, Families England.

Ferrer, G. (2006) Educational Assessment Systems in Latin America: Current Practice and Future Challenges. Washington-DC: *PREAL*. at URL: http://siteresources.worldbank.org/INTINDIA/4371432-1194542322423/21542197/LatinAmericanAssessmentSystems.pdf (viewed 28/04/14).

Fitz-Gibbon, C. (1991) Multi level modelling in an indicator system. In S. Raudenbush and J. D Willms (Eds.) *School, pupils and classrooms. International studies of schooling from a Multilevel perspective*. london and New York. Academic Press.

Gamboa, L. , Casas, A; Piñeros, L. (2003). La Teoría del Valor Agregado: una Aproximación a la Calidad de la Educación en Colombia. *Revista de Economía del Rosario*. 6 (2) pp. 95-116.

Gamboa, L. and Waltenberg, F. (2012) Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006–2009 *Economics of Education Review* 31 (2012) 694– 708.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D. and Thomas, S. (1993) A multilevel Analysis of School examination results. *Oxford Review of Education*. 19 (4): pp. 425-33.

Goldstein, H. (1997): Methods in School Effectiveness Research, *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 8:4, 369-395.

Gray, J., Reynolds, D., Fitz-Gibbon, C. and Jesson, D. (eds) (1996) *Merging Traditions: The Future of Research on School Effectiveness and School Improvement*, London: Cassell.

Gray, J, Goldstein, H and Thomas, S. (2003) Of Trends and Trajectories: searching for patterns in school improvement, *British Educational Research Journal*, 29/1, pp. 83-88.

Hanushek, E.; Peterson and Woessman, L. (2012) Achievement Growth: International and U.S. State Trends in Student Performance. at URL: http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG12-03_CatchingUp.pdf (viewed 28/04/14).Harker, R. and Tymms, P. (2004) The Effects of Student Composition on School Outcomes. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 15:2, 177-199.

Hoyle, R. and Robinson, J. (2003) League tables and school effectiveness: a mathematical model. *Proceedings of the Royal Society of London* B, 270, pp.113–199.

QAE (2013) Resumen Metodología de Ordenación. at URL: http://www.agenciaeducacion.cl/resumen-metodologia-de-ordenacion/ (viewed 1/05/14).

Lambert, D. and Lines, D. (2000). *Understanding Assessment. Purposes, perceptions, practice*. London: Routledge.

Ladd, H. and Walsh, R. (2002) Implementing value-added measures of school effectiveness: getting the incentives right *Economics of Education Review* 21, pp. 1-17.

Ladd, H. and Zelli, A. (2002) School-Based Accountability in North Carolina: The Responses of School Principals. *Educational Administration Quarterly*, 38, pp. 494-529.

Lareau, A. (2003) *Unequal Childhoods: Class, race and family life*. Berkeley: University of California Press.

Leckie, G. and Goldstein, H. (2007) *The limitations of using school league tables into inform school choice*. Ppt. presentation. at URL: http://209.85.229.132/search?q=cache:bR0DoCO0iaUJ:eprints.ncrm.ac.uk/265/1/School_league_tables.pdf+unfair+league+tables+filetype:pdfandcd=5andhl=enandct=clnkandgl=uk (viewed 23/03/2010).

Leckie, G (2008) *Multilevel Modelling of school differences in educational achievement*. Ph. D. thesis. Bristol: University of Bristol.

Leckie, G and Goldstein, H. (2011). Understanding Uncertainty in School League Tables. *Fiscal studies*, vol. 32, no. 2, pp. 207-224.

Luyten, H. (2003). The size of school effects compared to teacher effects: An overview of the research literature. School Effectiveness and School Improvement, 14, 31–51.OECD, 2008.

Manzi, J. San Martín, E. and Van Bellegem, S. (2011) *School System Evaluation by Value-Added Analysis under Endogeneity*. at URL: http://www.ecore.be/DPs/dp_1284546165.pdf 2010/90 (viewed 12/04/12).

Marsh, H., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A., Abduljabbar, A. & Köller, O. (2012): Classroom Climate and Contextual Effects: Conceptual and Methodological

Issues in the Evaluation of Group-Level Effects, Educational Psychologist, 47:2, 106-124.

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: An International Journal of Research, Policy and Practice*, 23:3, 305-326.

Mizala, A. and Romaguera, P. (2003). *Equity and Educational Performance.* Center for Applied Economics, Department of Industrial Engineering, University of Chile. at URL: http://www.webmanager.cl/prontus_cea/cea_2002/site/asocfile/ASOCFILE120030327125022.pdf (viewed 30/03/12).

Mizala, A., Romaguera, P. and Ostoic, C (2005). *Equity and Achievement in the Chilean School Choice System*. Center for Applied Economics, Department of Industrial Engineering, University of Chile. at URL: http://simce.cl/fileadmin/publicaciones-BD-simce/ASOCFILE120040507153638.pdf (viewed 30/03/12).

Mortimore, P. (1991) School Effectiveness Research: Which Way at the Crossroads? *School Effectiveness and School Improvement.* Volume 2, Issue 3, pp. 213-229.

Muñoz-Chereau, B. (2013) *Value added measures of school performance in Chile*. Ph. D. thesis. Bristol: University of Bristol.

Murillo, J., and Roman, M. (2011) The School or the Cradle? Evidences of its Contribution to Students' Performance in Latin America. Multilevel Study about the estimate of School Effects. *Profesorado.* VOL. 15, Nº 3, Diciembre, pp.26-50.

OECD (2008) *Measuring Improvements in Learning Outcomes. BEST PRACTICES TO ASSESS THE VALUE-ADDED OF SCHOOLS*. at URL: www.oecd.org/publishing/corrigenda (viewed 12/04/11).

OECD (2010) PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II). Paris: OECD Publishing. at URL: www.oecd.org (viewed 31/01/13).

OECD (2011a), Society at a Glance –OECD Social Indicators, Key Findings: Chile. at URL: http://www.oecd.org/dataoecd/39/23/47572883.pdf (viewed 12/06/12).

OECD (2013) PISA 2012 Results: Excellence through Equity Giving Every Student the Chance to Succeed.Volume II http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-II.pdf (viewed 31/01/15)

Opdenakker, M., & Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: An illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. School Effectiveness and School Improvement, 11, 103–130.

Opdenakker, M. & Van Damme, J. (2006) Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools. *School Effectiveness and School Improvement*, 17:1, pp. 87 - 117.

Ramírez, M. (2003). The distribution of mathematics knowledge among Chilean fourth graders and related explanatory factors. at URL: http:// 200.68.0.6/medios/20030604142903.pdf. (viewed 11/03/13).

Ramírez, M. (2007). Diferencias dentro de las salas de clases: Distribución del rendimiento en Matemáticas [Differences inside classrooms: the distribution of math scores]. Estudios Políticos, 106(5–22). at URL: http://www.cepchile.cl/dms/lang_2/ doc_3917.html. (viewed 11/03/13).

Rasbash, J., Jones, K., Steele, F. & Pillinger, R. (2009) Introduction to Multilevel Modelling in MLwiN. Practical Excercises. Center for Multilevel Modelling, University of Bristol. at URL: http://www.bristol.ac.uk/cmm/learning/course.html (viewed 11/03/13).

Raudenbush, S. and Willms, J. (1995) The Estimation of School Effects. *Journal of Educational and Behavioural Statistics*. December 21, vol. 20 no. 4 pp. 307-335.
Ray, A. (2006) *School Value Added Measures in England*. A paper for the OECD Project on the Development of Value-Added Models in Education Systems. Department for Education and Skills. at URL: www.teachers.tv/publications/eOrderingDownload/RW85.pdf (viewed 11/03/13).
Reardon, S. (2012) *Income, inequality, and educational outcomes: U.S. and international evidence*. Keynote at the 3rd meeting EARLI SIG. at http://www.earli.uzh.ch/programme/keynote-speeches (viewed 4/02/13).
Reimers, F. (2003) The social context of educational evaluation in Latin America. In T. Kellaghan and D. Stufflebeam (Eds.). *International Handbook of Educational Evaluation*. USA: Kluwer Academic Press. pp. 441-463.
Rodríguez-Jiménez, O. and Murillo, J. (2011) Estimación del efecto escuela para Colombia*. Magis, Revista Internacional de Investigación en Educación*, vol. 3, núm. 6, enero-junio, pp. 299-316.
Rosenkvist, M. A. (2010), Using Student Test Results for Accountability and Improvement: A Literature Review. *OECD Education Working Papers*, No. 54, OECD Publishing. at URL: http://dx.doi.org/10.1787/5km4htwzbv30-en (viewed 19/05/11).
Rowe, Ken (2003) "The Importance of Teacher Quality As A Key Determinant of Students' Experiences and Outcomes of Schooling" (2003). http://research.acer.edu.au/research_conference_2003/3
Sammons, P., Thomas, S., Mortimore, P., Owen, Ch. and Pennell, H. (1993) Assessing School Effectiveness: Developing measures to put school performance in context. *International School Effectiveness and Improvement Centre*. Institute of Education, London.
Sammons, P., Thomas, S. and Mortimore, P. (1997) Forging links: Effective Schools and Effective Departments. London: Paul Chapman Publishing.
Second Regional Comparative and Explanatory Study (SERCE) (2006) Results from Student achievement in Latin America and the Caribbean. at URL: http://unesdoc.unesco.org/images/0016/001610/161045E.pdf. (viewed 4/02/13).
Second Regional Comparative and Explanatory Study (SERCE-UNESCO) (2008). Segundo Estudio Regional Comparativo y Explicativo. Los aprendizajes de los estudiantes en América Latina y El Caribe. Reporte técnico. At URL: http://unesdoc.unesco.org/images/0019/001902/190297s.pdf. (Viewed 28/04/14).
Saunders, L. (2000). Understanding schools' use of 'value added' data: the psychology and sociology of numbers. *Research Papers in Education* 15(3) 2000, pp. 241-258.
Sammons, P., Cuttance, P., Nuttall, D. and Thomas, S. (1995) Continuity of School Effects: a longitudinal study on primary and secondary school effects on GCSE performance. *School Effects and School Improvement* 6 (4): pp. 285-307.
Sammons, P. (2007) *School Effectiveness and Equity: Making Connections*. A review of school effectiveness and improvement research. Its implications for practitioners and policy makers.
Sammons, P., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Hunt, S. (2008). Effective Pre-school and Primary Education 3-11 Project (EPPE 3-11): Influences on children's attainment and progress in Key Stage 2: Cognitive outcomes in Year 6. At URL: http://dera.ioe.ac.uk/18190/1/DCSF-RR048.pdf. (Viewed 28/04/14).
San Martín, E. and Carrasco, A. (2012) Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa. Temas de Agenda Pública, Centro de Políticas Públicas Uc. Año 7 , No 53 , julio, pp.1-18.
Shagen, I. and Hutchison, D. (2003) Adding value to educational research: the marriage of data and analytical power, *British Educational Research Journal*, 29 (5), pp. 749-765.
Scherrer, J. (2011) Measuring Teaching Using Value-Added Modelling: The Imperfect Panacea. *NASSP Bulletin* 95(2) pp.122–140.
Stoll, L. and Mortimore, P. (1997) School effectiveness and school improvement. In *Perspectives on School effectiveness and school improvement*. (Ed.) J. White and M. Barber. London: Institute of Education.
Slee, R., and Weiner, G. and Tomlinson, S. (1998) *School effectiveness for whom? Challenges to the school effectiveness and school improvement movements*. London: Falmer Press.

Smith, D. and Tomlinson, S. (1989) The School Effect. A study of Multi-Racial Comprehensives. Policy Studies Institute: Exeter.

Stoll, L. and Mortimore, P. (1997) School effectiveness and school improvement. In *Perspectives on School effectiveness and school improvement*. (Ed.) J. White and M. Barber. London: Institute of Education.

Thieme, C., Tortosa-Ausina, E., Prior, D. and Gempp. R. (2012). *Valor Agregado Multinivel y factores contextuales en educación: una comparación no paramétrica robusta*. at URL: https://editorialexpress.com/cgibin/conference/download.cgi?db_name=xveep&paper_id=53 (viewed 04/02/13).

Thomas, S and Mortimore, P. (1996) Comparison of Value Added Models for Secondary School Effectiveness, *Research Papers in Education*, 11(1), pp. 5-33.

Thomas, S. (2001): Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions, School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 12:3, pp. 285-322.

Thomas, S., Salim, M., Muñoz-Chereau, B. and Peng, W-J. (2012) Educational Quality, Effectiveness and Evaluation: Perspectives from China, South America and Africa, in Ch. Chapman, P. Armstrong, A. Harris, D. Muijs, D. Reynolds and P. Sammons (Eds.), *School Effectiveness and School Improvement Research, Policy and Practice. Challenging the Orthodoxy?*. London, Routledge pp. 125-145.

Thomas, S, Peng, W-J, HS Tian, JZ Li, & J Feng (2013) Evidence of CVA School Effects in China and UK. Paper presented at Symposium on Using Contextualised Value Added to examine school effectiveness as a tool for accountability. ICSEI 2013, Santiago Chile

Teddlie, C., and Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer Press.

Timmermans, A., C., Doolaard, S. and de Wolf, I (2011): Conceptual and empirical differences among various value-added models for accountability, *School Effectiveness and School Improvement*, 22:4, pp. 393-413.

Timmermans, A., C., & Thomas, S.M. (2014)). The impact of student composition on schools' value-added performance: a comparison of seven empirical studies. *School Effectiveness & School Improvement*. 10.1080/09243453.2014.957328

Treviño, E. and Donoso, F. (2010) Agrupación de escuelas para intervenciones de política: análisis del caso chileno at URL:http://mt.educarchile.cl/MT/jjbrunner/archives/Trevino_ValorAgregado2010.pdf (viewed 30/01/13).

Tymms, P. (1993) Accountability-can it be fair? *Oxford review of education* 19, 291-299.

Van den Noortgate, W., Opdenakker, M., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. School Effectiveness and School Improvement, 16, 281–303.

Willms, D., and Raudenbush, S. (1989) A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability. Journal of Educational Measurement, Vol. 26, No. 3 Autumn, pp. 209-232.