

Value-added indicators for a fairer Chilean school accountability system: A pending subject

Authors details

Author 1 (corresponding author)

Full name: Bernardita Muñoz-Chereau

Affiliation:

Senior Research Fellow in Education, Department of Learning and Leadership, UCL Institute of Education, University College London

Postal address: Room 808, 20 Bedford Way, Bloomsbury, London WC1H 0AL

Telephone: +44 (0)20 7911 5319

e-mail: b.munozchereau@ucl.ac.uk

Honorary Research Associate, Graduate School of Education, University of Bristol

email address: bernardita.munozchereau@bristol.ac.uk; edbmc@bristol.ac.uk

Full name: Bernardita Muñoz-Chereau

Author 2

Full name: Andrés Anwandter

Affiliation:

Honorary Research Associate, Graduate School of Education, University of Bristol

Postal address: 8 St. Michael's hill, Bristol BS28DT

Telephone: +44 (0)1172390013

email address: aanwandter@yahoo.com

Author 3

Full name: Sally Thomas

Affiliation:

Professor of Education, School of Education, University of Bristol

Postal address: Office 3.02, Helen Wodehouse Building, 35 Berkeley Square, Clifton
Bristol BS8 1JA

Telephone: +44 (0) 117 331 4382

email address: s.thomas@bristol.ac.uk

Word count: 6,244

Abstract

Although schools' relative contribution to pupils' progress is increasingly used in accountability systems around the world (OECD, 2013), momentum for value-added models (VAM) has not been reached in Chile. This small-scale study explores qualitatively the policy context in which this omission takes place, by analyzing policy documents and interviewing local policymakers and researchers about their views on VAM. These agents and official documents not only point to political, methodological and pragmatic reasons for and against value-added indicators, but also highlighted ethical and legal reasons pro (but not against) VAM. Overall, the most prominent reason for including VAM into the new accountability system was the ethical consideration. The notion that a fairer indicator could make justice, especially for those schools working in disadvantaged contexts, was a recurrent idea. In contrast, the most recurrent reasons against VAM were methodological. Whilst research on VAM for the Chilean school system has been conducted over the last decade, it has impacted very little the policy arena. Given that the creation of frameworks for assessing value-added indicators takes time, needs policy and research support and needs to consider potential unintended consequences, the road towards including them into the Chilean school accountability system is just starting.

Key words

Value-added, value-added models, Chilean educational system, school accountability, Chilean school system

Searching for fairer ways of assessing school effectiveness

Despite extensive literature over 25+ years addressing this topic there remains an ongoing international debate on how to best assess the contribution that schools make to pupils' learning. Notably few researchers claim that raw-score results (or unadjusted tests) are the best way for assessing school effectiveness because it matters very little which school a child goes to (Gorard, 2006, 2008, 2010). Contrary to Gorard's claim, educational effectiveness research conducted in different contexts has shown that schools have indeed a critical effect upon children's academic

achievement and progress, demonstrated by statistically significant differences between schools in relation to pupils' value added progress. Whilst raw measures of pupils' academic outcomes are useful for highlighting the attainment gaps at the school system, it is typically considered a poor indicator for other purposes -such as school accountability, school improvement or school choice- because making the school solely responsible for the results of their students, without accounting for external factors, is not fair, nor defensible (Muñoz-Chereau & Thomas, 2016).

Taking a cross-sectional approach at one time point, contextualized attainment models control for student-level socio-economic and demographic factors without a longitudinal dimension that accounts for measures of prior attainment. Using this alternative approach, research has shown that measures produced by these models, although better than raw attainment scores, do not remove satisfactorily the effects of previous attainment background characteristics, especially in unequal school systems (McCaffrey, Lockwood, Koretz & Hamilton, 2004). As reported by Sammons, Thomas, Mortimore, Owen and Pennell (1993) as well as Thomas and Mortimore (1996), when prior attainment is absent, socioeconomic factors provide an imperfect proxy to explain a substantial proportion of the variation in student outcomes. Thus the lack of individual prior attainment results in substantial increases in the apparent impact of some of the control variables, such as parental education. Also, in the absence of prior attainment, the overall goodness of fit of the models tends to be low. For example Sammons et al. (1993) reported a goodness of fit in terms of only 13.3% of the total variance explained, after controlling for student-level characteristics/background and school-level context variables simultaneously when analyzing a sample of schools' GCSE (General Certificate of Secondary Education) performance in England, whereas models including prior attainment typically explain up to 50% of total variance (Thomas & Mortimore, 1996).

Also approaches such as regression-discontinuity designs (RDD) widely used to analyse the intervention of other specific programmes using Chilean administrative data, appears not feasible to assess yearly school performance. A reason for this was provided by Urquiola and Verhoogen (2009) when evaluating the effect of class size on student academic outcomes. These authors

reported that a key assumption of RDD –that schools cannot manipulate their treatment status– was violated, as Chilean schools implemented strategic behaviours (e.g increasing fees as a way of reducing enrolments) instead of opening new classrooms to tackle oversubscription whilst keeping class size constant. Therefore RDD internal validity can be compromised when implemented in Chilean context.

The position that researchers and policymakers take within the debate concerning the best approaches to evaluate school effectiveness is crucial, especially in countries with a high-stake test-based accountability system such as England and Chile, where certain school communities, headteachers and teachers are rewarded, ignored or penalized on the basis of how effective they appear to be. Therefore this paper explores school value added models (VAM) which estimate the relative contribution that schools make to their students' educational progress towards prescribed education objectives, whilst controlling for the contribution of other factors outside the control of the school that help explaining that progress (OECD 2008). Although the international literature has yet to reach consensus regarding the value of VAM, the fact that raw and contextualised measures routinely used in the Chilean accountability system are highly correlated with students' SES, makes it difficult to identify high value-added schools (McEwan, Urquiola, Vegas, Fernandes and Gallego, 2008). Although VAM are far from being a panacea, and have not developed without controversy, they have been conceived as a more accurate tool than raw results for identifying good practice in the educational system (OECD, 2008). In the Chilean context, local researchers have recognised VAM as a sound methodological alternative for analysing the educational impact of market-driven mechanisms in education (Carrasco and San Martín, 2012), and critically evaluating the effectiveness of the school system, different educational policies, and the new accountability system in place (Page, San Martín, Orellana and González, 2016). Consequently, the advantages and shortcomings of VAM identified in the literature will be explored in the next section.

VAM strengths

Accuracy: Comparing like with like

Due to the fact that pupils and teachers are not distributed randomly in most school settings, individual student characteristics (also called "the intake") are considered to be potentially

confounded with those associated with classes or school groups (Raudenbush and Bryk, 2002). This has proved especially true in the Chilean educational system (Manzi, San Martín and Van Bellegem 2011; Meckes and Bascopé, 2012; Torres, 2018). The accuracy argument has been presented as follows: "natural justice demands that schools are held accountable only for those things they can influence (for good or ill) and not for all the existing differences between their intakes" (Nuttall, 1990, p.25). The principle applied here is that "like should be compared with like" (Gray and Wilcox, 1995, p.89). The OECD (2008) also highlighted that VA is a technique designed to make fair comparisons between schools by giving a more accurate measurement of their performance: "It yields estimates of average progress for each institution. To assess the "value" added by the school, it is essential to adjust for various background factors and for prior attainment by the individual child. To assess effectiveness of different schools without taking such information into account is like comparing apples with oranges. Unless schools are compared on a like with like basis, judgements are neither fair nor valid" (Stoll and Mortimore, 1997, p.10; Ballou, Sanders and Wright, 2004).

VAM have been presented as an alternative and valid tool for making comparisons in school performance. After testing out different ways of considering students' differences in SES and their prior cognitive capacity, "the adoption of value added methodology enables the calculation of more reliable estimates of school relative effectiveness (residuals) in promoting students' outcomes. Studies emphasize the importance of obtaining individual student level information about personal, family and socioeconomic background and, of crucial importance, about prior attainment, in the analysis of school effects on cognitive achievements" (Sammons, 1999, p.56).

Even though there is a general agreement around this principle of the need of comparing like with like, authors are still inconclusive on how to implement it. What factors are "exogenous" or outside school control? Or, to put it differently, what should schools be held accountable for? The overarching rationale behind VAM is that it is possible to adjust those factors outside the control of the school that can be identified as partially responsible for school differences. This claim, which lies at the heart of this field, is rarely questioned. But being aware that "education is jointly produced by teachers, schools, families, and communities" (Hanushek, 1997; Harris, 2009, in Rosenkvist, 2010, p. 27), the idea is that "the use of such models is intended to emulate

(to the greatest extent possible) the situation of a randomised experiment" (OECD, 2008, p.156). As defined by the influential work of Raudenbush and Willms (1995), "school effect" may be the extent to which attending a particular school modifies a student's outcome" (p. 308). So, in order to tackle the problem of non-random assignment of students to schools, VAM adjust for pre-existing differences among students and schools.

Sophistication

VAM are sophisticated when compared with earlier assessment models within School Effectiveness Research (SER) using different approaches (Thomas, Peng and Gray, 2007; Braun, Chudowsky and Koenig, 2010), often requiring the mediation of technical aspects before their results are used at the school level (Saunders, 2000; Rosenkvist, 2010). Most sophisticated statistical techniques currently available have enabled a better understanding of the relative importance of the contribution of the school to student outcomes (Goldstein, 1995).

Caution in interpretation: Let's be realistic

VAM need to be interpreted cautiously since confidence intervals for school effects indicate many schools cannot be reliably separated one from another in terms of their estimated contribution to students learning. Given that most of the schools will have overlapping confidence intervals, especially when single cohorts are examined, residuals cannot be used as rankings and only a few schools' future performances can be separated from both the overall mean and from one another with an acceptable degree of precision (Leckie and Goldstein, 2009). However, they can be classified in three groups: underperforming, average or over-performing schools (Goldstein and Thomas, 1996). For this reason, studies in this area need to take note of the statistical uncertainty attached to estimates of individual school effects (i.e. 95% confidence interval). Only then can VAM be used as helpful screening instruments (Sammons, 1999). So, if a school is below expectations -across several criteria and over a period of years- then legitimate concerns can be raised about its performance. Rather than being passive or complacent, VAM means being realistic.

VAM limitations

VAM weaknesses evolve around the claims of unreliable estimations in small schools, low expectations, questionable utility of estimates, complexity and endogeneity.

Unreliable estimations in small schools

Even though shrunken estimates - resulting variances that are smaller than the true values obtained in the estimates of regression predictions- (see Goldstein, Rasbash, Yang, Woodhouse, Pan, Nuttall and Thomas, 1993) generate a more precise estimation when weighting the importance of cluster units according to their sizes, precision and stability of the estimates become problematic when estimating the performance of small schools with few students enrolled. To solve this issue, previous studies conducted in Chile have excluded from their samples schools with less than a certain number of students, such as 20 or 15 (Carrasco and San Martín, 2012), which imposes a limitation in assessing small schools, particularly rural ones.

Low expectations

One of the side effects of VAM is the strong evidence that the impact of background or contextual factors on value-added measures can lead to lower expectations for specific groups of students (Sammons, 1999). As it is metaphorically expressed by Braun et al (2010, p. 8) "Value-added modelling can make the playing field more level, but it can also reverse the tilt". Timmermans. et al (2011) problematized this point by saying that "using ethnicity or gender as a control variable implies that we expect and allow some subgroups of students to perform less successfully on their final examination, regardless of their prior achievement" (p. 419).

Questionable utility of estimates

VAM will make a difference to a school's reported performance only at the extremes of effectiveness and ineffectiveness (OECD, 2008). Thus it has been argued that some sectors -such as prospective employers- are not interested in the students' qualifications adjusted by their social background, but instead on what their educational outcomes show they can do (Schagen and Hutchison, 2003). Therefore, it is important to compare unadjusted measures of school effect with VAM because even though the latter are fairer to assess school performance, they can also hide differences in Raw scores. VAM can only refer to the sample studied, and what is more,

estimates of school performance are affected by the choice of which control/explanatory variable is included.

Complexity

As recognised by many authors, there is a trade-off between presenting a more easily communicable model and developing a model that is more statistically robust but also more complex (Saunders, 2000; OECD, 2008; Evans, 2009; Manzi, San Martín and Van Belleghem, 2011). Complexity derived from the statistics on which VAM unavoidably rest represents a challenge for potential users when interpreting the findings. The most fundamental objections are concerned with what the adjustment processes do. This is especially problematic given the evidence that "a new system is likely to be more effective if teachers believe the measure treats them fairly in the sense of holding them accountable for things that are under their control" (Braun et al, 2010, p. 22) which, of course, needs to be understood in the first place. In this line, interesting work oriented to communicate VA results to schools for improvement purposes was conducted by the National Foundation for Educational Research (NFER) when these models became part of the accountability system in England (i.e. Schagen and Morrison, 1999; Saunders, 2000). Their work evidenced the difficulties of communicating complex information on school performance to stakeholders to encourage school accountability and improved performance.

Endogeneity

One major difficulty that VAM face, is the problem derived from the endogeneity of school inputs (i.e. economic resources) due to the non-random way in which they are allocated across schools, especially in the Chilean case. Given that economic resources and other omitted variables are already correlated with both the actual test score and the prior test score, the error of the model is not uncorrelated with the prior attainment score. The implication of this non-vanishing correlation between the prior attainment score of a student and the school random effect is that the estimation of the school random effect will be affected by this correlation (Manzi, San Martín and Van Belleghem, 2011). Therefore these authors have included instrumental variables to account for this problem.

Summarizing, the three major VAM strengths are: firstly, the accuracy argument that highlights fairer comparisons between school performance after adjusting for factors that are out of the school's control (such as background factors and prior attainment). Secondly, the VAM research designs, techniques and software for data analysis are more sophisticated when compared with earlier models within School Effectiveness Research (SER)/ Educational Effectiveness Research (EER) using different approaches, making clear that it is foolish to pretend that a Raw league table can fairly assess the quality of schools. Thirdly, VAM calls for a realistic interpretation of school effects because confidence intervals for VAM estimates are wide, meaning that many of the schools cannot be reliably separated one from another in terms of their estimated contribution to students' learning. The reason why this is seen as an advantage is because VAM need to be interpreted with a degree of caution.

Despite their merits, these approaches are not free from caveats, nor do they guarantee they will serve all policy purposes. As Braun et al (2010) summed it up: "Each major class of models has shortcomings, there is no consensus on the best approaches, and little work has been done on synthesizing the best aspects of each approach (...) More needs to be learned about how these properties differ, using different value-added techniques and under different conditions" (p. 54). It is recognised that the main limitations mentioned above -unreliable estimations in small schools, low expectations for specific group of students, questionable utility for specific stakeholders, complexity and the problems derived from the endogeneity of schools' inputs- have a strong impact on these models' merits. Moreover, in those educational systems where VAM has been implemented to evaluate teacher effectiveness such as the USA, weaknesses and unintended consequences have been argued to be stronger, as researchers have raised additional significant limitations such as the risks of test score inflation (Hursh, 2013; Koretz, 2017). As summarised by Koretz (2017), "it's utterly irrational to "evaluate" teachers based on scores earned by students of other teachers, particularly teachers in other schools or who teach other subjects" (p.145).

The context: Chilean new accountability system

After three decades of neoliberal reforms, the Chilean educational system stands out for the wrong reasons: low academic performance, high socioeconomic segregation and inequality of

educational opportunities (Valenzuela, Bellei & De Los Ríos, 2014). This system has been singled out internationally as one where students' socio-economic status typically results in one of the highest influences on students' performance (OECD, 2010). As clearly put by Murillo & Román (2011), the cradle has an overwhelmingly stronger effect than the school when explaining Chilean pupil academic performance. Additionally, Chile is among those countries such as Macao-China, Hong Kong-China, the Netherlands and Ireland, where fewer than 50% 15-year-old students attend public schools (OECD, 2013). Because Chilean schools are also unequally funded, making fair and accurate comparisons between schools is very difficult.

In this context, Chile has introduced new school accountability mechanisms, which are coherent with a performative turn in its educational system (Pitton, 2012). The “Sistema Nacional de Aseguramiento de la Calidad de la Educación” (National System of Education Quality Assurance) encompasses the creation of different bodies, such as the Quality Agency of Education (QAE), whose role is to monitor and assess school performance in order to make schools accountable for their results. Recently the QAE has used contextualized attainment models and other non-academic indicators of school quality (such as students' citizenship and self-esteem) in order to classify schools in four performance groups: High, Medium, Low Medium and Insufficient. High-stakes consequences on the basis of this classification - ranging from organizational interventions up to school closure – are implemented for schools that do not meet the required standards (i.e. the 630 “insufficient” schools).

As this new classification does not use VAM to account for Chile's extremely unequal society, we argue that, although moving from raw league tables to contextualized attainment models represents an improvement, nevertheless, the new system of school classification is unfair and still punishes schools located in disadvantaged areas. Not surprisingly, according to the 2017 QAE classification, only 3.3% of schools with High and Middle-High Socioeconomic Status are “Insufficient”, in comparison with 31.4% of those with Low and Middle-Low Socioeconomic Status. By contrast, 74.8% of schools with High and Middle-High Socioeconomic Status are classified as “High” performing, in comparison to 18.2% of those with Low and Middle-Low Socioeconomic Status (QAE, 2017).

Recent research using Chilean data has concluded that the apparent differences in school effectiveness reduce dramatically when using VAM. Muñoz-Chereau (2018) found that the percentage of total variance in student's math Raw attainment attributable to differences between primary schools reduced from 28.5% to 14.5% after controlling for aspects arguably out of the school control, indicating that the pupil intake and contextual features of Chilean primary schools exert a powerful influence on students' outcomes. Without VAM, these differences might be wrongly attributable to differences between schools' effectiveness, when in reality they are by large due to prior attainment and socioeconomic inequalities (Manzi, San Martín & Van Belleghem, 2011; Muñoz-Chereau, 2013; Muñoz-Chereau & Thomas, 2016; Torres, 2018). Whilst in England the correlations between raw and value-added models tend to be relatively high, in the Chilean case these tend to be relatively low. Here lies another key reason to include these types of indicators into the Chilean accountability system. Crucially Muñoz-Chereau (2013) found a correlation of 0.31 between Math raw and contextual value-added residuals in secondary schools, and Carrasco and San Martín (2012) reported a similar low correlation of 0.42 between Math raw and VAM residuals in a sample of Chilean secondary schools. This differs significantly with research carried in England, where Thomas & Mortimore (1996) and Gorard (2008) found a much higher although still imperfect association between raw and value-added residuals: 0.73 and 0.75, in Math and English, respectively.

For these reasons, without value-added indicators, the new school classification system in Chile will say little about the quality of the schools and much about prior attainment of their intakes, seriously biasing the perception of the effectiveness of Chilean schools, a point strongly supported by previous researchers (McEwan & Carnoy, 2000; Bellei, 2005; Schagen & Hutchison, 2003; Hsieh & Urquiola, 2006; Carnoy, 2007; Valenzuela, 2008; San Martín & Carrasco, 2012). In line with Murphy (2012), we are not arguing for value-added as the only indicators to be considered when implementing high-stakes decisions. Rather, and in line with Saunders and Rudd (1999), we advocate for the need to include these indicators among others, in order to construct "a truer picture of a school's achievements, and for diagnosing pedagogical strengths and weaknesses within the school" (p. 9), which are the expected effects that these measures can bring if they were eventually included in the Chilean school accountability system.

Although value-added indicators appear specially fit for complementing the estimation of Chilean schools' effects, and different initiatives have promoted their inclusion in the new accountability system, these indicators are absent from the current educational landscape. Yet from the 2000s onwards, value-added indicators for school performance seem to be waiting for government ratification, as their inclusion has been repeatedly recommended in the Chilean context (Himmel report, 2003; LGE, 2011).

Thus, this paper seeks to elucidate the policy context in which value-added indicators remain disregarded and address the following research question: what do Chilean policy documents and stakeholders report as the main advantages and barriers of using VAM to evaluate schools?

Research Design and Methodology

The research is conducted by analyzing policy documents and interviewing local policymakers and researchers about their views on VAM. We will explore why and how the new assessment system keeps reflecting only the Socioeconomic Status (SES) of the students but not their relative progress. The aim is to seek explanations that enable us to understand why, despite a favourable legal context towards the consideration of value-added indicators, their inclusion into the new school accountability system has been held back.

Materials and Methods

Policy Documents

Six official policy documents that define the new national assessment system were included in this study. These were identified by one expert informant at the Curriculum and Assessment Unit (Ministry of Education), and confirmed via an independent literature search by the authors as the key documents emanated from government agencies that have shaped the current national assessment system:

1. Himmel report: SIMCE report commissioned by the Minister of Education and lead by Erika Himmel, 2003

2. LGE: Creates the National Quality Assurance System for Pre-school, Primary and Secondary Education and its control. Official Magazine of the Republic of Chile, Santiago, August 27, 2011
3. Recommendations for a National Curriculum Development Policy, May 2016
4. Response NAP: Response to the Recommendations of the National Council of Education Regarding the National Assessment Plan 2016-2020, Ministry of Education, June 2016
5. NAP: National Assessment Plan (2016-2020)
6. Decree 182, that established the National and International Assessment Plan 2016-2020, Official Magazine of the Republic of Chile, Santiago June 20th, 2016

Interviews

Five semi-structured interviews with key policymakers (one current and one former officer from QAE) and researchers that have collaborated with QAE and work in different universities (known here as University A, B and C for anonymity reasons) were conducted in Chile during March and April, and October and November 2017.

The aim of the analysis was to answer the research question oriented to identify the reasons given by policy documents, policymakers and researchers in favour and against using value-added indicators to assess school performance. After translating from Spanish to English and transcribing the policy documents and interviews, we took a grounded approach to inductively build an interpretation based on constant comparisons. Instead of analyzing each source separately, we identified component themes or categories across policy documents and interviews (Kohler Riessman, 2008). This procedure consists in breaking down into smaller pieces the information and compare the pieces for similarities and differences before grouping them under themes or categories (Corbin and Strauss, 2015). In this way the themes or categories were derived from data during the analysis. By careful consideration of the data, an interpretation of the most common meanings was brought into the analysis. Five different types of justifications (legal, ethical, methodological, political and pragmatic) and component themes (i.e. fairness, social justice, etc.) were identified by the researchers. In this sense, the justifications and themes were constructed by the researchers based on the interpretation of themes and categories.

Results

The inclusion of VAM to measure the academic performance of Chilean schools presents multiple possibilities and challenges. In terms of the reasons supporting the integration of VAM into the new accountability system, five different types of justifications provided by policy documents, policy-makers and researchers were identified.

Table 1: Reasons pro value-added provided by policy makers, researchers and policy documents

Reasons	Pro VAM	Exemplary quote
Legal	Inclusion in LGE and Decree 182	“In order to carry out this arrangement, the Agency shall consider the learning results of all the areas measured in the national assessments census survey, the distribution of the pupils' results in relation to the learning standards and the degree of achievement of the other indicators of educational quality. It shall also consider the school pupils' characteristics, including, among others, its vulnerability, and, where appropriate, progress or value-added indicators. Nevertheless, gradually, the arrangement of the schools will tend to be carried out independently of the socioeconomic characteristics of the pupils, to the extent that the system corrects the differences ascribable to those characteristics in their performance.” (LGE, 2011, Art. 17 N° 3)
Ethical	Fairness	“The Commission considers relevant that the system allows to know the learning progression throughout the school process and not only the status in each level or evaluated course. In other words, record the student's performance in a level or course in relation to their performance in the following or earlier year. To achieve this, it is imperative that the annual sequence of measurements allow the same cohort to be followed up. This can be achieved by ensuring that the cohort that participates in the measurement in 4th grade is evaluated in 8th grade and, desirably, also in 10th grade. It highlights the importance that this information may have for the incentive system and for the public account of schools, in terms of doing fairer assessments, which reflect not only the performance of the students, but also their progress. On the other hand, a modality of this type would contribute to discourage selection and exclusion practices by schools, by explicitly assessing progress with respect to different starting points.” (Himmel Report, 2003, p. 61)

	Social justice	<p>“When you isolate the variables, above all socioeconomic level, you can see that in a four year period the public schools, in relative terms, grow much more [...] even though they have lower SIMCEs, but they start much lower [...] whereas some types of schools, above all the fully private, keep a high SIMCE, but the history of their SIMCE is always high, so we don't know the leaps they have made [...] and also since there is already a lot of knowledge pointing out that SIMCE only explains socioeconomic differences, then I believe that making a value-added assessment would do justice to schools, could also improve the collective self-esteem of the schools, of this teacher that goes to a bad school, because everybody is telling him it has 190 points, but nobody looks at the history, the trajectory of that school [...] I think it would introduce a bit more justice.” (Researcher University A)</p>
Methodological	A new perspective/ report status and progress	<p>“I think it would be very interesting to bring up the idea that value-added indicators could substitute other indicators of quality [...] Thus, it would make me happy if the solution to that problem {lack of validity of new indicators} were longitudinal value-added. That you report status and progress, and the other indicators of quality, such as citizenship and academic self-esteem.” (Former policymaker QAE).</p>
	Reliable indicator	<p>“I think that value-added indicators yield much more information than values that are [...] cross-sectional basically, that take different cohorts, always in the same moment, because the cohorts are different.” (Researcher University B)</p>
Political	International policy borrowing	<p>“I think that we'll be late, as always [...] when someone comes out with it, when a new minister arrives [...] probably, if this government is elected, they will go to England or the United States. And will come back with the idea that they, in Mississippi they do this value-added thing...oh yes, it rings a bell, we did it last government, it didn't work, let's do it again, and we will be the kings of value-added [...] with a five year delay in respect to the developed world.” (Former Policymaker QAE)</p>
	Regional Leadership	<p>“In Latin America there is no other country with a working value-added system (In Higher Education) in Colombia, and I am not sure, they are like just starting [...] but none in school education, no other country. Thus Chile would become the first.” (Researcher University C)</p>

Pragmatic	Feasible given the availability of longitudinal data	“This Plan allows the estimation of some progress or value-added indicators, because it makes possible for some student cohorts to be assessed twice during their school life. For example, in some cases, we will be able to do a follow-up between year four and six, and in others, between year eight and year ten.” (NAP, p. 34)
-----------	--	---

First of all, documents and participants mentioned (1) legal reasons, understood as claims referred to frameworks or documents developed by law, such as the inclusion in LGE and Decree 182. For example “The Evaluation Plan shall allow the carrying out of studies of progress and/or value-added indicators, in line with the recommendation of the Committee that reviewed SIMCE in 2003 and as established in the law N° 20.529.” (Decree 182, p. 12). In this way the LGE (2011) made explicit the need to conduct value-added studies associated with the new classification of schools used for the implementation of high-stakes accountability measures. (2) Ethical reasons grouped claims referring to principles and ideals that should be promoted in the educational system, such a making the school system fairer or socially just. In line with Braun et al (2010), who pointed out the ethical and methodological reasons that support value-added indicators in the U.S context, one of the strongest reasons given by policymakers and researchers (and also present in policy documents) was bringing fairness - understood as fair-play achieved through the consideration of the context in which schools are located - and social justice to the current school classification system. As expressed by a researcher “I think it is a good idea that we look for information that is more sensitive towards the context, and hence I see that value-added should be a positive addition, even if the law had not mandated it.” (Researcher University C). (3) Methodological reasons refer to considerations of the procedures, methods and issues that need to be conducted or resolved before generating value-added indicators, such as data, sample and models. Regarding methodological reasons, three interviewees perceived that value-added models imply a new perspective, enabling the reporting of status alongside progress measures, and gaining a reliable indicator to disentangle the school effect. For example, “A value-added indicator should [...] provide a perspective and a kind of information that differs from what we've been traditionally looking at. So, I see it as an opportunity to enrich the system.” (Researcher University C). (4) Political reasons point out to government agendas and administration priorities. The political reasons mentioned (international policy borrowing and regional leadership) suggest that local policymakers and researchers are aware and subject to the

influence of regional and international trends in educational policy. Finally, (5) Pragmatic reasons group practical considerations, such as the opportunity to conduct value-added studies because there is available data. Overall, the most prominent reason for including value-added indicators into the new accountability system was the ethical consideration. Given the national and international recognition of the unequal educational playing field present in Chile, the notion that a more accurate and fairer indicator could make justice, specially to those schools working in disadvantaged contexts, was a recurrent idea among participants and policy documents. By contrast, the pragmatic reason appeared as the less prominent.

Regarding reasons against value-added indicators, although policy documents, policymakers and researchers identify more challenges than advantages associated with VAM, these can be grouped in only three groups of reasons: Political, Methodological and Pragmatic. In other words, no legal or ethical considerations against value-added were identified in this study.

Table 2: Reasons against value-added provided by policy makers, researchers and policy documents

Reasons	Against VAM	Exemplary quote
Political	Value-laden methodology	“The right-wing wanted to close schools, it didn't matter which ones, and the left-wing wanted to avoid closing public schools. [...] That was the truth. And hence, given the extreme segregation [...] multilevel allowed you to estimate that you would be less harsh with those little public schools.” (Former policymaker ACE)
	Lack of political vision	“Because I think that for this administration value-added would have been more palatable than what they have today. And there, well, political decisions are never taken for the country, always for the most immediate interest, but if the QAE's committee, which was pro-government in its majority, had decided to implement the methodology suggested by the opposition, that was multilevel, it is much more probable that they would have kept it when in office [...] there was a lack of political vision, of knowing that this was stillborn.” (Former policymaker ACE).
	Promotes competition between schools	“This system is much more aggressive in modelling improvement, thus requiring a lot of emphasis in schools' orientation and support. And this leads you to a certain logic, that is the logic of cooperation, of building networks of schools, so that they can improve by cooperating [...] the

		state cooperates and also trusts [...] this logic is very strong in this system of quality assurance, but it coexists with the logic of accountability, with the consequence of closure, which is like [...] a logic much more of distrust, a more punitive logic, more pressure, more stick than carrot [...] this law makes these two logics coexist [...] we have emphasized and invested much more in the first one, and in this sense value-added log behind [...] because it goes against the grain.” (Policymaker ACE)
	Research opposition	“They never dare to... and there it can be helpful to consider the national climate [...] the last year of Piñera's government, when they were implementing it, was the first year of the QAE, there was a lot of hostility against this issue, [...] a strong aggressiveness within academic circles against this [...] the idea that this was a stupidity, badly done, too simplistic, and that it had been possible to implement only because they had the political majority in the QAE's committee. [...] there was such an opposition towards ordinary least squares, that it was too difficult to even start discussing value-added” (Former policymaker ACE)
Methodological	Narrow focus on academic outcomes	“It generates a lot of pressure in schools and goes against the principles of this administration, of rationalizing the number of tests, of widening the perspective over quality, of stop putting all the assessment effort on academic performance tests. I think that, in technical terms, there are thousands of reasons that support doing value-added models. But politically it is a different issue.” (Policymaker ACE)
	Lack of vertical equating	"In order to do it in an adequate manner you have to ensure that the test can be compared between the different years, and in Chile SIMCE is not comparable, because it is not vertically equated.” (Researcher, University B); “We also have problems linked to the metrics of the scales, that is [...] the system is working at different levels, but the tests for each level are not connected between them or aligned, they have formally equivalent scales but [...]they are not vertically equated.” (Researcher University C)
	Debatable indicator	“Furthermore, there is some awareness, or vague awareness, that the concept of value-added is at least debatable or controversial. That it is not a sort of panacea as some believed it would be.” (Researcher University C)
	Biased sample	“Those schools for which it is plausible to estimate progress and value-added indicators are characterized by particular conditions, for example, high enrolment, high student retention, they are predominantly fully private or

		state-subsidized private [...] they represent a biased sample of the schools' population, limiting the external validity of the estimations carried out for monitoring the educational system.” (NAP, 2016, p. 34)
	School size	“There are many special issues that are difficult to sort out in Chile. One of them is the issue of the schools' size, that is, we have many small-sized schools. And how do we solve the problem of value-added when schools have too little information?” (Researcher University C)
	Low consistency	“(There is) evidence and debate within the academic community regarding the low consistency of value-added indicators estimated with different methodologies and control variables. Because of this, the results are very sensitive to technical decisions along the estimation process.” (NAP, 2016, p. 34) “We will have to monitor that, and to see how consistent or inconsistent it is [...] between years, between disciplines, throughout the years, [...] all that can be solved, but not all of those solutions are well settled yet, for example, if you want to observe value-added for a longer perspective [...]not just looking at lots of two but bigger blocks, it is not obvious how [...] the missing set is amplified, and you need more assumptions for the statistical models, becoming very complicated.” (Researcher University C)
	High mobility of students	“The problem of establishing value-added methodologies at census level is that there is too much mobility of schools in Chile. [...] 30% of the students move from one school to another [...] because from one year to another you have 30% of mobility, you have to do value-added in one year, and that is almost impossible.” (Policymaker QAE)
Pragmatic	Schools overloaded with external assessment	“Now the problem with value-added indicators is that they require a greater number of tests. And one of the big issues we have faced is that we see that schools are too overloaded with externalized assessment, generating some tension within them. And to generate tension within schools through policy is not good, because it prevents public policy from working as it wants to work. Thus, in fact, one of the things that this administration did was to rationalize the number of tests [...] we went from 18 to 9, that is, we cut the tests by a half.” (Policymaker ACE)
	Limited longitudinal data	“With the evaluations plan established last year, we have given up to have any type of value-added [...] there is less availability of data on the pupils' trajectory.” (Former policymaker QAE)

	Long periods between assessments	“For example, you could follow a student from year two, to year four, to year eight in primary, and to year ten in secondary. You wouldn't have SIMCE for all those years, but you could have it at least for year four and eight. But these levels have a four year difference, and this is too much. In general, one should have one every two years or so, but the idea is not to flood them with tests.” (Researcher, University B)
	Long-term project	“If you opened the value-added box, you wouldn't get it (the new school classification for accountability purposes) because you had to give it much more thought, you had to try, you had to pilot, you had new instruments, you had to see if it made any sense, you had to check the stability of the indicators [...] it was an impossible deadline, and a technical standard that looked too difficult” (Former policymaker ACE) “Implementation was too quick (the new methodology), without going through a wider discussion and obtaining empirical evidence about how it would work, before adopting one model [...] apart from a study that lasted a couple of months.” (Researcher University B)
	Limited applicability	“It is important to point out that, on one hand, given the high mobility of students between different year groups in the Chilean educational system and the huge proportion of small schools in year four, it is not possible to generate progress or value-added measure for all Chilean schools, limiting its applicability for accountability at the school level.” (NAP, p. 34)
	Expensive	“Implies increasing the amount of years [...] the cost of doing this is too high [...] too costly for an administration that needs to take charge of each and every school. Thus, it is very difficult in practice [...] it is too difficult to implement [...] I think the main problem is that you have to increase the number of tests. And that implies resources.” (Policymaker QAE)

In terms of the (1) Political reasons, participants argued that VAM were value-laden concepts, associated with test-based accountability, that tend to align better with right-wing governments educational reforms. They also mentioned the lack of political vision of previous administrations, as the development of value-added indicators takes longer than the 4 years period in which government is in power. It was highlighted that the current focus of the government was to promote collaboration instead than competition between schools, which was identified as a by-product of VAM. Policy makers argued that they had encountered such research opposition

against the current system, that thinking in a more complex one was beyond their possibilities. They also mentioned that VAM promoted a narrow focus on academic outcomes, something that was identified as a big limitation of the new accountability system. In this sense policymakers were more interested in evaluating new no-academic outcomes (such as self-esteem), rather than bringing more complexity into the way academic outcomes were assessed. In terms of (2) Methodological reasons against VAM, the lack of experts in the country was limiting its implementation, as well as the lack of vertical equating in SIMCE tests, and the debatable nature of the value-added indicator. Also the biased sample, resulting from the presence of many small schools in which the value-added indicator cannot be estimated, was perceived as a limitation: “Those schools for which it is plausible to estimate progress and value-added indicators are characterized by particular conditions, for example, high enrolment, high student retention, they are predominantly fully private or state-subsidized private [...] they represent a biased sample of the schools' population, limiting the external validity of the estimations carried out for monitoring the educational system.” (NAP, 2016, p. 34). Another methodological limitation was the low consistency that the value-added indicator reach when estimated with different methodologies and control variables, and the high mobility of students. As stated by a policymaker: “The problem of establishing value-added methodologies at census level is that there is too much student mobility between schools in Chile. [...] 30% of the students move from one school to another [...] because from one year to another you have 30% of mobility, you have to do value-added in one year, and that is almost impossible.” (Policymaker QAE). Focusing on (3) Pragmatic reasons, participants mentioned that schools were already overloaded with external standardised assessments, so the limited longitudinal data was a real problem, just as the long periods between assessments (i.e. 4 years) that characterise the national assessment system. Another pragmatic issue was the limited applicability of the indicator, as for example its problematic application in rural small schools. Finally, the development of new tests on which to base VAM was perceived as too long and expensive, especially given the fact that VAM was out of the policy priorities currently under implementation.

Discussion

This small-scale study explored qualitatively the policy context in which VAM have been omitted from the Chilean educational landscape, by analyzing policy documents and

interviewing local policymakers and researchers about their views. Firstly, although participants and official documents point to political, methodological and pragmatic reasons for and against value-added indicators, they only highlighted ethical and legal reasons pro (but not against) VAM. Aligning the ethical dimension with Lakoff (2002), who argued that “it is the work of the government to “level the playing field” for the disadvantaged” (p. 180), at a time of growing pressure and associated vulnerability derived from the introduction of high-stakes consequences on schools -particularly those classified as ‘insufficient’- there is local awareness that VAM could level the playing field particularly for those schools that are not showing high standards on attainment, to be rightly recognized for the value they are adding to their pupils’ progress. By providing performance indicators that do not measure students’ relative progress over time, QAE will continue conveying the unsubstantiated messages (i.e. school type explains student performance, public schools are inefficient), contributing to the vulnerability of public schools that may not be in a position to demonstrate, due to influential but unmeasured factors outside their control, high standards on raw scores, as well as supporting undesirable consequences, such as school selectivity, segregation, unfair competition and low school morale in more disadvantage contexts. Concerning the legal dimension, although incorporating VAM into the new school accountability system is prescribed by law, neither policymakers nor researchers perceive it as urgent or imperative. Rather they understand it as a possibility, among others, for improving the system.

Secondly, despite participants and policy documents provided a more restricted set of reasons against VAM, many political, methodological and pragmatic barriers to including them into the new accountability system were identified. Although some of the difficulties claimed (i.e high mobility of students and small schools) are common to any attempt of using statistical models to measure school effectiveness, they also affect VAM and need to be carefully considered before implementing these models in the Chilean educational system. However, some of the perceived limitations are open to discussion. For example, the idea that it is unadvisable to track students for long periods of time is not supported by previous research. The Chicago Public school test-based accountability system is well-known for tracking students’ outcomes up to four years (Jacob, 2005). Indeed, longer time periods between prior and outcome attainment are likely to result in more variation in student attainment and more robust VAM.

Without dismissing the challenges identified in this study, and in line with Murphy (2012), we are not arguing for value-added as the only indicators to be considered when implementing high-stakes decisions. Rather, along with Carrasco and San Martín (2012) and Saunders and Budd (1999), we advocate for the need to include these indicators among others, in order to construct “a truer picture of a school’s achievements, and for diagnosing pedagogical strengths and weaknesses within the school” (p. 9), which are the expected effects that these measures can bring if they were eventually included in the Chilean school accountability system.

We believe that the debate around VAM for school performance should move in two directions. The first one is to critically review the local research into longitudinal value-added indicators that has been conducted using Chilean performance data over the last decade. Our previous work -along with the research conducted by leading researchers who have explored various models and indicators of value-added using SIMCE data- will provide a clearer picture of the possibilities and challenges that need to be resolved before moving from research to policy. Although there is no consensus regarding the VAM that should be specifically adopted to account for the Chilean case, critical engagement with the available evidence could enhance the role that value-added can play in a fairer assessment of the effectiveness of Chilean schools. Secondly, as the methodological and pragmatic challenges could be resolved, the most critical one seem political. Whilst the previous administration chose a methodology for classifying schools that appeared more efficient and avoided using multi-level modelling because it was perceived as being against their overall policy agenda (i.e. implement a high-stake accountability system rather than develop a fair system), the current government shifted away from value-added indicators as a way of prioritizing a more supportive role of QAE, away from testing and competition. Because without political will it is unlikely that VAM will be implemented in the Chilean context, and given that the political reasons were not just used to justify the decision of excluding them from the national policy, but also the need to be “cautious” when approaching studies taking this approach (Decree 182, p. 13), it is clear that the policy discussion is heavily loaded against these indicators: they are seen as potentially detrimental to the accountability system currently in place, based on contextualized attainment models and other indicators of school quality. Bearing in mind that the creation of high-quality frameworks for assessing value-added indicators takes time, needs support (Gray & Wilcox, 1995; Saunders, 1997), and needs to

consider limitations and potential unintended consequences (Goldstein & Spiegelhalter, 1996; Gorard, 2006, 2008, 2010; Hursh, 2013, Koretz, 2017), the road towards including them into the Chilean school accountability system is just starting. Moving from being a research tool to an instrument for judging school quality will require accessible examples of how these indicators can be used at the national and school level to monitor students' progress, alongside other types of evaluation evidence, and support school improvement efforts. Until then, value-added indicators will keep being omitted from the Chilean school accountability system.

Finally, bearing in mind that this is a small-scale qualitative study, it is important to highlight the limitations derived from its methodology. The study does not have an explanatory status and it is not representative of a larger population. The limitations derived from working with an opportunistic sample are well known. However, its descriptive nature has given us a window to identify the main strengths and limitations articulated by key policy documents, policymakers and researchers around VAM in the Chilean context, which in turn helps to clarify and understand the key issues that need to be addressed before seeing this type of indicator being used at the national level.

References

Agencia de la calidad/QAE (2017). Agencia de la Calidad entregó categorías de desempeño a todos los establecimientos de Chile. [Quality Agency delivered performance categories to all Chilean schools] Retrieved from

<http://www.agenciaeducacion.cl/noticias/agencia-calidad-entrego-categoria-desempeno-todos-los-establecimientos-chile>

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Educational and Behavioural Statistics*, 29, 1, 29-37.

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

from <http://www.periglobal.org/sites/periglobal.org/files/BELLEI%20-PEPG-05-13%20Private%20Public%20in%20Chile.pdf>

Braun, H., Chudowsky, N., & Koenig, J. (2010). *Getting value out of value-added*. Washington D.C: The National Academies Retrieved from <http://rlhoover.people.yasu.edu/ClassConnections/core-readings/Getting%20Value%20Out%20of%20Value-Added.pdf>

Carnoy, M. (2007). Improving quality and equity in Latin American education: A realistic assessment. *Pensamiento Educativo*, 40, 1, 103-130.

Carrasco, A., & San Martín, E. (2012). Voucher system and school effectiveness: Reassessing school performance difference and parental choice decision-making. *Estudios de Economía*, 39(2), 123-141. Retrieved from <http://econ.uchile.cl/uploads/publicacion/783e3b83492db94c151bac5a7aa7e0c4d368671a.pdf>

Corbin, J. and Strauss, A. (2015). *Basics of qualitative research. Techniques and Procedures for Developing Grounded Theory*. (4th Edition) Sage: London.

Decree 182 (2016) Retrieved from <https://www.leychile.cl/Navegar?idNorma=1094644&buscar=Establece+Plan+Nacional+de+Evaluaciones+Nacionales+e+Internacionales+para+el+Per%ADodo+2016++2020>

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

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. (1995). *Multilevel statistical models*. London: Edward Arnold.

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)*, doi:10.2307/2983325.

Goldstein, H and Thomas, S. (1996) Using Examination Results as Indicators of School and College Performance, *Journal of the Royal Statistical Society*, 159(1), (pp. 149-163).

Gorard, S. (2006). Value-added is of little value, *Journal of Education Policy*, 21, 2, 235-243.

Gorard, S. (2008). The value-added of primary schools: what is it really measuring?, *Educational Review*, 60, 2, 179-185. Doi: 10.1080/00131910801934185

Gorard, S. (2010). Serious doubts about school effectiveness, *British Educational Research Journal*, 36, 5, 745-766.

Gray, J., & Wilcox, B. (1995). *Good school, Bad school*. Buckingham: OUP.

Himmel Report (2003) *Evaluación de Aprendizajes para una Educación de Calidad*. [Learning Assessment for a Quality Education] Retrieved from [http://www.facso.uchile.cl/psicologia/epe/_documentos/GT_cultura_escolar_politica_educativa/recursos%20bibliograficos/articulos%20relacionados/comisionsimce\(2003\)informe.pdf](http://www.facso.uchile.cl/psicologia/epe/_documentos/GT_cultura_escolar_politica_educativa/recursos%20bibliograficos/articulos%20relacionados/comisionsimce(2003)informe.pdf)

Hsieh, Ch-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90, 8-9, 1477-1503.

Hursh, D. (2013) Raising the stakes: high-stakes testing and the attack on public education in New York, *Journal of Education Policy*, 28:5, 574-588,

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of public Economics*, 89, 5-6, 761-796.

Kohler Riessman, C. (2008). *Narrative methods for the human sciences*. Sage: London.

Koretz, D. (2017). *The testing charade: Pretending to make schools better*. Chicago, IL: University of Chicago Press.

Lakoff, G. (2002). *Moral politics: How conservatives and liberals think*. Chicago: U of Chicago.

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of Royal Statistic Society* 172, 4, 835-851.

LGE [Ley General de Educación], (2011). Retrieved from <https://www.leychile.cl/Navegar?idNorma=1006043>

Manzi, J. San Martín, E., & Van Belleghem, S. (2011) *School System Evaluation by Value-Added Analysis under Endogeneity*. Retrieved from http://www.ecore.be/DPs/dp_1284546165.pdf 2010/90

McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2004). Models for Value-Added Modelling of Teacher Effects. *Journal of Educational and Behavioural Statistics*. Spring, 29, 1, 67-101.

McEwan, P., & Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis* 22, 3, 213-39.

McEwan, P. J., Urquiola, M., Vegas, E., Fernandes, R., & Gallego, F. A. (2008). School choice, stratification, and information on school performance: Lessons from Chile. *Economia*, 8(2), 1-42.

Meckes, L., & Bascopé, M. (2012). Uneven Distribution of Novice Teachers in the Chilean Primary School System. *Education Policy Analysis Archives*, 20, 30, 1-27

Muñoz-Chereau, B. (2013). *Searching for fairer ways of comparing Chilean secondary schools performance: a mixed methods study investigating contextual value added approaches* (Doctoral dissertation, University of Bristol).

Muñoz-Chereau, B., & Thomas, S. M. (2016). Educational effectiveness in Chilean secondary education: comparing different 'value added' approaches to evaluate schools. *Assessment in Education: Principles, Policy & Practice*, 23(1), 26-52.

Muñoz-Chereau, B. (2018). Exploring gender gap and school differential effects in mathematics in Chilean primary schools. *School Effectiveness and School Improvement*, 1-21.

Murillo, J., & Román, 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*, 15, 3, 26-50.

Murphy, D. (2012). Where is the value in value-added modeling. 24–26 September 2014, Nanyang Polytechnic, Singapore.

NAP [National Assessment Plan 2016-2020] Retrieved from <https://www.curriculumenlineamineduc.cl/605/w3-article-34980.html>

Nuttall, D. (1990). *Differences in examination performance*. London: Research & Statistics Branch, ILEA. RS 1277/90.

Organisation for Economic Co-operation and Development (2008). Measuring Improvements in Learning Outcomes. Best practices to assess the value-added of schools. Retrieved from www.oecd.org/publishing/corrigenda

Organisation for Economic Co-operation and Development (2010). PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II). Paris: OECD Publishing. Retrieved from www.oecd.org

Organisation for Economic Co-operation and Development (2013)/ School Governance, Assessments and Accountability. Chapter 4. In What Makes Schools Successful? Resources, Policies and Practices. Volume IV. Retrieved from <https://www.oecd.org/pisa/keyfindings/Vol4Ch4.pdf>

Page, G. L., Martín, E. S., Orellana, J., & González, J. (2016). Exploring complete school effectiveness via quantile value added. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 315-340.

Pitton, V. (2012) Neoliberalization of educational policy discourses in Chile. (Doctoral dissertation, University of Illinois at Urbana-Champaign). Retrieved from <http://hdl.handle.net/2142/32033>

Raudenbush, S. W., & Willms, J. (1995). The estimation of school effects. *Journal of educational and behavioral statistics*, 20(4), 307-335.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd. Edition). Sage: London.

Rosenkvist, M. A. (2010). Using Student Test Results for Accountability and Improvement: A Literature Review. *OECD Education Working Papers*, No. 54, OECD Publishing. Retrieved from URL: <http://dx.doi.org/10.1787/5km4htwzvbv30-en>

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. (1999) *School Effectiveness coming of Age in the Twenty-first Century*. Swets and Zetlinger: The Netherlands.

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 institutions: contribution of value-added models for a fair accountability] *Temas de Agenda Pública*, Centro de Políticas Publicas UC. 7, 53, 1-18.

Schagen, I. and Morrison, J. (1999). A methodology for judging departmental performance within schools, *Educational Research*, 41:1, pp. 3-10.

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.

Saunders, L. (1997). Value-added principles, practice and ethical considerations. In A. Harris , N. Bennett and Preedy, M. (Eds), *Organizational Effectiveness and Improvement in Education*. (pp.191-202). Buckingham: Open University Press.

Saunders, L., & Rudd, P. (1999) Schools' use of value-added data: a science in the service of an art? Paper presented at the British Educational Research Association Conference, University of Sussex, at Brighton, 2-5 Sept., 1999.

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.

Shagen, I., & Hutchison, D. (2003). Adding value to educational research: the marriage of data and analytical power, *British Educational Research Journal*, 29,5, 749-765.

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.

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., Peng, W.J. and Gray, J. (2007). Modelling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, 33(3), pp. 261–295.

Timmermans, A., C., 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.

Torres, R. (2018). Tackling inequality? Teacher effects and the socioeconomic gap in educational achievement. Evidence from Chile. *School Effectiveness and School Improvement*, DOI: 10.1080/09243453.2018.1443143.

Urquiola, M., & Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99(1), 179-215.

Valenzuela, J. (2008). Segregación en el Sistema Escolar Chileno: en la búsqueda de una educación de calidad en un contexto de extrema desigualdad. In P. Velasco (Ed.), *Transformaciones del espacio público*. (pp. 131-156). Santiago: Universidad de Chile.

Valenzuela, J. P., Bellei, C., & De Los Ríos, D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, 29,2, 217-241.