Fundación BBVA

Documentos de Trabajo

5

2013

Claudio Thieme
Diego Prior
Emili Tortosa-Ausina
René Gempp

Value Added and Contextual Factors in Education

Evidence from Chilean Schools

Fundación BBVA

Value Added and Contextual Factors in Education

Evidence from Chilean Schools

Claudio Thieme[1]
Diego Prior[2]
Emili Tortosa-Ausina[3,4]
René Gempp[1]

[1] *DIEGO PORTALES UNIVERSITY*
[2] *AUTONOMOUS UNIVERSITY OF BARCELONA*
[3] *UNIVERSITY JAUME I*
[4] *THE VALENCIAN INSTITUTE OF ECONOMIC RESEARCH (IVIE)*

Abstract

There is consensus in the literature about the need to control for socioeconomic status and other contextual variables at student and school level in the estimation of value added models, for which methodologies rely on hierarchical linear models. However, this approach is problematic because the nature of their estimate is a comparison with a school mean, implying no real incentive for performance excellence. Meanwhile, activity analysis models recently developed to estimate school value added have been unable to control for contextual variables. We propose a robust frontier model to estimate contextual value added which integrates recent advances in the activity analysis literature. We provide an application to a sample of schools in Chile, where reforms have been made in the educational system focusing on the need for accountability measures. Results indicate the general relevance of including contextual variables, and explain the performance differentials found for the three school types.

Resumen

Existe consenso en la literatura sobre la necesidad de controlar el nivel socioeconómico y otras variables contextuales a nivel de escuela y/o estudiante en la estimación de modelos de valor añadido, para el que las metodologías se basan en modelos lineales jerárquicos. Sin embargo, este enfoque es problemático debido a que esta estimación se establece a partir de una comparación con la media de la escuela, lo que no supone incentivo alguno de cara a la excelencia. Mientras tanto, los modelos de análisis de actividad desarrollados recientemente para estimar el valor añadido escuela no han sido capaces de controlar las variables contextuales. En este documento se propone un modelo de frontera robusta para estimar el valor añadido contextual que integra los últimos avances de la literatura sobre análisis de actividad. Se ofrece una aplicación para una muestra de las escuelas chilenas, donde se han realizado reformas en su sistema educativo destacando la necesidad de medidas de rendición de cuentas. Los resultados indican la importancia general de la inclusión de las variables contextuales, y explican los diferenciales de rendimiento encontrados para los tres tipos de escuelas.

Key words

Efficiency, order-*m*, school effectiveness, value added.

Palabras clave

Eficiencia, orden-*m*, eficacia escolar, valor añadido.

Al publicar el presente documento de trabajo, la Fundación BBVA no asume responsabilidad alguna sobre su contenido ni sobre la inclusión en el mismo de documentos o información complementaria facilitada por los autores.

The BBVA Foundation's decision to publish this working paper does not imply any responsibility for its contents, or for the inclusion therein of any supplementary documents or information facilitated by the authors.

La serie Documentos de Trabajo tiene como objetivo la rápida difusión de los resultados del trabajo de investigación entre los especialistas de esa área, para promover así el intercambio de ideas y el debate académico. Cualquier comentario sobre sus contenidos será bien recibido y debe hacerse llegar directamente a los autores, cuyos datos de contacto aparecen en la *Nota sobre los autores*.

The Working Papers series is intended to disseminate research findings rapidly among specialists in the field concerned, in order to encourage the exchange of ideas and academic debate. Comments on this paper would be welcome and should be sent direct to the authors at the addresses provided in the About the authors *section.*

La serie Documentos de Trabajo, así como información sobre otras publicaciones de la Fundación BBVA, pueden consultarse en: **http://www.fbbva.es**

The Working Papers series, as well as information on other BBVA Foundation publications, can be found at: **http://www.fbbva.es**

Versión: Octubre 2013
© Claudio Thieme, Diego Prior, Emili Tortosa-Ausina and René Gempp, 2013
© de esta edición / *of this edition:* Fundación BBVA, 2013

EDITA / PUBLISHED BY
Fundación BBVA, 2013
Plaza de San Nicolás, 4. 48005 Bilbao

1. Introduction

THE development of indicators to evaluate the quality of education is a core element of countries' efforts to implement improvements in their education systems (Battauz et al. 2011). In particular, in the field of public policies on education, there is a growing concern about the evaluation of students (Denvir and Brown 1986; Ercikan 2006).

In many countries, this concern has motivated the adoption of accountability systems (Kane and Staiger 2002), whose main objective is to evaluate school quality and report these results to parents, principals, teachers, or policy makers, who will use them to make choices about schools, to improve their professional practice or to develop educational policies[1]. The available empirical evidence in this regard has contributed to strengthen this tendency, showing that well designed accountability systems (i.e. those which find the responsibility attributable to each of the participants in the educational system) enable organizational improvement inside each school (Rouse et al. 2007), as well as optimizing the educational outcomes (Carnoy and Loeb 2002; Hanushek and Raymond 2005).

Usually, accountability systems model school quality as the educational achievement of the students in those schools, assuming that schools are responsible for the largest share of their students' academic achievement. Therefore, an underlying requisite of any accountability system is to use a robust methodology to disentangle what share of the students' achievement can be attributed to the school, and what share is simply the result of the student's motivation, abilities, socioeconomic capital, or other variables beyond the school's control.

In terms of methodology, the general consensus is that students' educational achievement depends both on their personal characteristics and those of their school and context. In order to analyze these scenarios, the most common and accepted methodology is multilevel regression models (McCaffrey et al. 2004), also known as hierarchical linear models, or regression models with random effects (Goldstein 2003; Raudenbush and Bryk 2002). The key characteristic of these methods is their capacity to disentangle what proportion of variance in student achievement can be explained by student variables (level 1), and what share can be explained by aggregate, or school, contextual variables (level 2). When multiple levels are

[1] By way of example, see, for instance, the detailed information regarding school performance in the UK disclosed in http://www.education.gov.uk/schools/performance/.

considered, such as hierarchical systems of students nested in schools, it is possible to obtain a better understanding and measurement of the causes that explain students' learning processes (Aitkin and Longford 1986). The multilevel approach is highly relevant when attempting to make decisions, specific to each student, school, or context, that contribute useful information to develop new improvement processes in schools, discourage managers' opportunistic behavior, signal a correct resource endowments policy (by establishing rewards and penalties), and make decisions on public policies.

In this regard, there is a growing consensus about the kind of student achievement measure needed to build valid indicators of school performance. Although the initial stages of research on school accountability were characterized by the use of cross sectional measures to estimate school performance (e.g. the mean annual results of standardized tests), the current practice is to rely on panel data methods to evaluate student performance, in order to estimate the academic growth of students throughout their school life —ideally also controlling for other relevant variables (Goldstein and Thomas 1996; Goldstein et al. 1993; Gray et al. 1995; Mortimore et al. 1994; Sammons 1995). In fact, there is a consensus among researchers that an effective school is not the one where students achieve the best results, but the one where students make greater progress than they would have made in another school with a similar background. In other words, effective schools are those which *add value* to their students' achievement compared to other schools serving student populations with equivalent characteristics.

In the context of school accountability, the value added (VA) of a given school can be broadly defined as the contribution that it makes to students' net progress (i.e. to the learning objectives) after the effects of other variables, external to the school, have been removed (Meyer 1997); in other words, the extent to which schools do make a difference (Coleman et al. 1966). In its most basic form, the estimation of value added requires a set of statistical procedures to analyze the longitudinal performance of a sample of students nested in schools, in order to make inferences about the contribution of each school to the academic growth of its students (Raudenbush 2004; Tekwe et al. 2004). Thus, the basic value added model compares schools' performance controlling for students' previous achievement.

More complex value added models are also available, and over the last few years there has been a growing tendency to use *contextual* VA models, which allow researchers to control for socioeconomic status (SES), ethnic background, gender, and other variables that are not under the school's control or responsibility. Thus, contextual VA models provide an estimation of the net performance of schools by removing the effect of previous achievement and other preexisting differences among students (Ballou et al. 2004). It is generally agreed that contextual variables should be used to estimate VA models, especially when setting some form of accountability, or when disseminating the results, since results might be questionable if they do not take into account contextual characteristics of both students and schools. Although there is no consensus as to what specific contextual variables should be included in the model (Tekwe et al. 2004), socioeconomic status (SES) is usually one of them.

Because VA models estimate the net contribution of the school to the educational growth of their students, they are useful to compare effectiveness across schools —even though their students' populations might be quite heterogeneous— and provide a guide for planning educational improvements, both at school and public policy levels (Drury and Doran 2003; McCaffrey et al. 2004). Moreover, VA indicators emerge as an attractive alternative for several actors interested in measuring or improving school performance, including: (i) governments (which need to rely on objective accountability measures); (ii) politicians (who want to guarantee that the assessment of schools considers their ethnic and socioeconomic diversity); (iii) researchers (who need to study those factors contributing to school effectiveness using net indicators, which are not spuriously contaminated by the characteristics of students); (iv) teachers and school managers (who want objective measures of their performance, tuned to their specific student populations); (v) parents (who need to choose schools for their children according to their real capacity to add value to their students); and (vi) society as a whole, since this entails a more accurate and fair evaluation of the schools in the country.

It is also crucial to understand that school effectiveness studies —including VA analysis— require using some kind of methodology to compare the schools being evaluated with a benchmark. In the case of VA research, the most popular methodology is multilevel regression models. Some examples of this approach are the studies by Goldstein et al. (1993), Gray et al. (1995), Cervini (2009), or Blanco (2010), among others. In this case, estimating VA models using regression analysis involves comparing the performance of any given school with the performance of the average school under analysis. Thus, an implicit assumption of the traditional approach to estimate VA in education is to use the average school as a benchmark. However, this approach is not free from criticisms, one of them being that using the average as a benchmark is not an incentive for excellence (Bock et al. 1996; Kupermintz 2003; McCaffrey et al. 2003).

An attractive approach to overcome this criticism is to consider the models derived from the activity analysis literature, which evaluate school performance by comparing any given school with the best observed performance. Instead of using a regression line as a benchmark, these methodologies consider a nonparametric frontier built either using data envelopment analysis (DEA), or its non-convex variant, namely, free disposal hull (FDH)[2]. In addition to explicitly defining an optimal benchmark, frontier models also allow several outputs to be used simultaneously (i.e. several concurrent measures of student and school performance), offering greater flexibility to estimate VA.

In this line of research, there has been a growing interest in developing approaches based on this literature to estimate not only basic VA models (see, for instance Silva Portela and Thanassoulis 2001; De Witte et al. 2010; Portela et al. 2012) but also to analyze contextual effects in multilevel settings (Thieme et al. 2013). However, the existing methodologies have not been able to estimate contextual VA, namely, to develop a frontier model able to estimate school VA effects controlling for students' previous achievement, and also for contextual variables at student and school levels. This development is crucial to further explore the use of frontier models to estimate contextual VA models in real world applications.

For this reason, the aim of this paper is both methodological and empirical. At the methodological level, we propose a *robust* frontier model to estimate contextual value added (CVA) which integrates both methodological contributions from multilevel modeling to school VA, as well as relatively recent proposals in the field of activity analysis methods—namely, the so-called *metafrontiers* (Battese et al. 2004) as well as the partial frontier methods (Cazals et al. 2002).

At the empirical level, we use this novel approach to analyze school effectiveness in Chile. This application is especially relevant for this country which, since the 1980s, has been implementing a series of reforms to its educational system (see Mizala and Romaguera 2000), with strong emphasis on accountability measures. Among other reforms, the government transferred the management of public schools from the Ministry of Education to city councils, and allowed for the participation of private schools in the public system through a voucher system.

[2] We can also find parametric variants to this literature, among which SFA (stochastic frontier analysis) is the most popular. Parametric and nonparametric methods have both advantages and disadvantages, some of which have been recently outlined by Badunenko et al. (2012).

Simultaneously, an accountability system was created, consisting of national standardized tests of educational achievement applied annually to all students in 4th, 8th or 10th grade. The average school results of this assessment, called the SIMCE test (*Sistema de Medición de la Calidad Educativa*, or Measurement System of Educational Quality), are reported to parents, and used by the Ministry of Education as a measure of school quality. Due to some particularities of the SIMCE test calendar, 2009 was the first time in which the same students coincided in two exams corresponding to different levels (4th and 8th grade).

This scenario allows us to apply our model to a large sample —47,076 students from 948 primary schools. All students took mathematics and language SIMCE tests. The sample was made up of 4th and 8th grade students (9 and 13 years old respectively), for whom we have socioeconomic information on their families (at student level), obtained via a questionnaire for parents. In an attempt to achieve a reliable and homogeneous sample at school level, we included only those students in schools who took both exams and for whom we had socioeconomic information, whose schools met the requirement of having more than 30 students meeting these criteria, and this value corresponded to 60% of the students where who took the exam in 2009.

Results can be explored in several dimensions. First, a relevant finding is that omitting contextual variables could result in managers blaming these centers for variables that are beyond their control, or to create perverse incentives to select students according to their socioeconomic characteristics. Second, we also find that performance differences among privately-owned fee paying, privately-owned subsidized, and public schools diminish substantially when controlling for contextual variables. Third, from an empirical point of view, the results suggest that most of the large differences observed in the results of academic achievement among types of schools are largely explained away when the variables that account for students' entry conditions and contextual variables are factored in. Therefore, much of the variance explained by the school, which is particularly high in Chile due to the high level of school segregation in the country, disappears once these variables enter the model.

The rest of the article is organized as follows. In Section 2 we describe the relevant theoretical framework. In Section 3 we detail the methodology used. The background of the empirical application and the description of the database used are presented in Section 4. The comparative results between models are discussed in Section 5, and the main conclusions of the study are outlined in Section 6.

2. Theoretical Framework

2.1. The assessment model

All national or state accountability systems attempt to improve learning and instruction processes, but they differ significantly in the way they control for both the quality and progress of schools. This heterogeneity leads to different perceptions of which schools should be rewarded, and which should be encouraged to improve, among other recommendations.

There are several frameworks to classify these evaluation models (Carlson 2001), but most of them take into account two fundamental aspects. First, one may distinguish between two different approaches for monitoring school performance, namely, *status* models and *growth* models. Status models use a single year to evaluate students' academic achievement (i.e. cross sectional data); whereas growth models use two or more years (i.e. panel data). Second, in both approaches one may distinguish between models that use contextual variables (both at student and school level) to evaluate school achievement, and those which do not. This scenario, and the core questions that these models try to answer, can be represented in a 2×2 matrix as in table 1.

TABLE 1: **Types of evaluation models**

	Without contextual variables	**With contextual variables**
Status (one student assessment)	*Model 0:* What is the level of academic achievement of the students in this school?	*Model 1:* Which is the level of academic achievement of the students in this school, according to the students and/or school contextual factors?
Value added (two or more students' assessments)	*Model 2:* Is this an effective school? According to the achievement of students upon enrolment, how much do they learn or develop while they are at school?	*Model 3:* Is the school more effective? Given students' achievement level upon entrance, how much do they learn, or develop, while they are at school, according to either the students' or school's contextual factors?

As table 1 shows, *Model 0* (which is also referred to as type 0 model) only considers for the evaluation the outputs related to students' academic achievement in a given time period. In the literature on education these models are usually referred to as academic achievement status models without contextual variables. As indicated by Tekwe et al. (2004), the distinguishing characteristic of status-based models is the absence of adjustment for students' incoming knowledge level. This would imply that the differences among schools in terms of the average knowledge of their incoming students are convoluted with the assessment of teaching quality. This scenario is relatively straightforward to model. An implicit assumption is that all students and schools have optimal and similar backgrounds. Therefore, accountability systems based on this model consider that students' academic achievement is entirely attributable to schools, disregarding evidence in the literature that a large share of students' academic achievement might be attributable to contextual factors, which are non-controllable, and not attributable to the school itself (Teddlie and Reynolds 2000).

The strongest criticisms suggest that this model could generate perverse incentives for the attainment of the objective being pursued, endowing fewer resources to those students with relatively worse results who do not help their schools to achieve their objectives. Simultaneously, this could generate selection of students within schools, or lead to self-selection (Wilson 2004).

In spite of these disadvantages, we analyze Model 0 as a first step, because it is the approach currently used in Chile to evaluate its schools and, therefore, it is of interest to compare its results with those that could be yielded by other models proposed in this study.

Model 1 extends the variables considered. While, analogously to model 0, it includes the outputs related to students' academic achievement at a given moment of time, it also considers input variables not attributable to the school, either at student or at school level. This model corresponds to an academic achievement status model *with* contextual variables, according to the literature on education. A recent example of this type of approach is the study by Thieme et al. (2013), which proposes a multilevel model incorporating contextual variables both at student and school levels. Despite the remarkable progress it represents, by not considering as input the initial academic achievement of students, it assumes that it is the same, and optimal, for all students, a situation which is obviously far from reality, and could lead to misinterpretation of results.

Model 2 corresponds to a *pure* value-added model; the only inputs and outputs it considers are the results of students' academic achievement, both at the beginning and at the end of the educational process under evaluation. The educational research literature considers that value added measures (gain) are more informative measures of the effectiveness of institutions, since they allow the effect of the school's student progress to be isolated (Wilson and Piebalga

2008), and contribute to reducing incentives for dishonest behavior. Two relatively recent studies are consistent with this model, namely, Silva Portela and Thanassoulis (2001) and De Witte et al. (2010). However, as in the previous model, they have the disadvantage of not controlling for non-school elements which influence this particular process.

Model 3 overcomes the disadvantages of models 1 and 2. In the education research literature this type of model emerged strongly as a refinement of measures of growth, and has been called CVA (contextual value added). The CVA was first used in 2006 in British schools, and it is a measure intended to isolate the *real* impact of the school on students' progress. This type of modeling involves obtaining results that consider a number of factors such as gender, ethnicity, and language of origin, among others. The difference between the model estimate and the result that the student actually achieves is what is referred to as CVA (Wilson and Piebalga 2008).

Despite the great advances that models 1 and 2 represent, when activity analysis methods are considered to evaluate them, model 3 in table 1 best isolates the real impact of the school on students' progress, as indicated by many contributions from the traditional literature on school evaluation —which generally use parametric multilevel analysis. Therefore, evaluating model 3 considering activity analysis methods has some unexplored advantages that will be part of our aims.

2.2. The evaluation methodology

As indicated in the introduction, in recent years there has been considerable progress in the evaluation methodology of school performance, especially regarding the development of multilevel models (Bryk and Raudenbush 1992; Goldstein 1995). The general concept is that students' academic achievement depends on their personal characteristics, and the characteristics of the school, and its context. To analyze these situations, the different levels are considered as hierarchical systems of students and schools, with individuals and groups defined in separate hierarchies, using variables that are defined at each level (Hox 2002).

This significant progress can solve the main methodological problem of the pioneering studies in this field, by breaking down the various nested effects that explain students' educational outcomes. The percentage of student achievement due to the different variables at different organizational levels —district, school, class, and student— can also be determined.

In this particular area, there are many statistical models for estimation which differ in several regards such as the definition and inclusion of adjustment variables (Tekwe et al. 2004). However, the most prominent position is to include adjustment variables, especially when establishing some form of accountability or dissemination of the results, since the equity is questionable if the background characteristics of students and schools are not taken into account (McCaffrey et al. 2003, 2004).

Despite the many methodological and empirical contributions, this research is not without its criticisms (Kupermintz 2003; McEwan 2003). One of them is related to the nature of their estimate as a comparison with the average, assuming no real incentive for performance excellence. Indeed, the vast majority of value-added studies used multilevel regression or analysis.

An alternative is found in the models that consider activity analysis techniques, mainly using non-parametric frontier methods (mostly DEA, and its non-convex counterpart, FDH). They provide relevant advantages such as the ability to compare with the optimal or, more importantly, the possibility to specify several inputs and outputs simultaneously. In the field of education many studies have adopted these techniques (see, for instance Bessent et al. 1982; Ruggiero et al. 1995; Mancebón and Mar Molinero 2000; Mizala et al. 2002; Ouellette and Vierstraete 2005, among others). However, these methods are not exempt from general criticisms. On the one hand, regarding the nature of these methods, their deterministic and probabilistic features, the curse of dimensionality, or their heavy reliance on the absence of outliers have been a source of continuous concern. On the other hand, in the specific field of education, their main disadvantage has been to consider only student-level data, which would yield estimations that incorrectly assume that schools are operating with the optimal endowment of inputs (both, controllable or uncontrollable), without establishing thus a multilevel analysis.

Both types of problems —i.e. those more intrinsic to the methods chosen, and those more intrinsic to the general problem under analysis— are addressed in this paper, by providing an integrated approach which merges contextualized multilevel analysis contributions from the field of education (McCaffrey et al. 2003, 2004), with some others from the field of activity analysis —in particular metafrontier approaches (Battese et al. 2004) and the use of partial frontiers (order-*m*) (Cazals et al. 2002).

Some previous research initiatives have taken these considerations into account. Specifically, our aims and methods are consistent with previous literature such as Silva Portela and

Thanassoulis (2001) who, following model 2, decompose the overall efficiency into two different effects, namely, school effect and student-within-school effect. Later on, De Witte et al. (2010) refined this methodology, proposing a robust approach based on Cazals et al.'s (2002) ideas for the estimation. A more recent contribution by Thieme et al. (2013), based on model 1, considered both ideas, performing a multilevel decomposition in which additional variables are factored in so as to provide a more comprehensive analysis (i.e. they consider not only the school effect but also a student-within-school effect, a resource endowments effect, and a peer effect, all of which have been proved to be relevant by the literature on education). However, despite their interest, these previous studies disregard the existence of contextual factors —at both the student and school level— in the assessment of school performance.

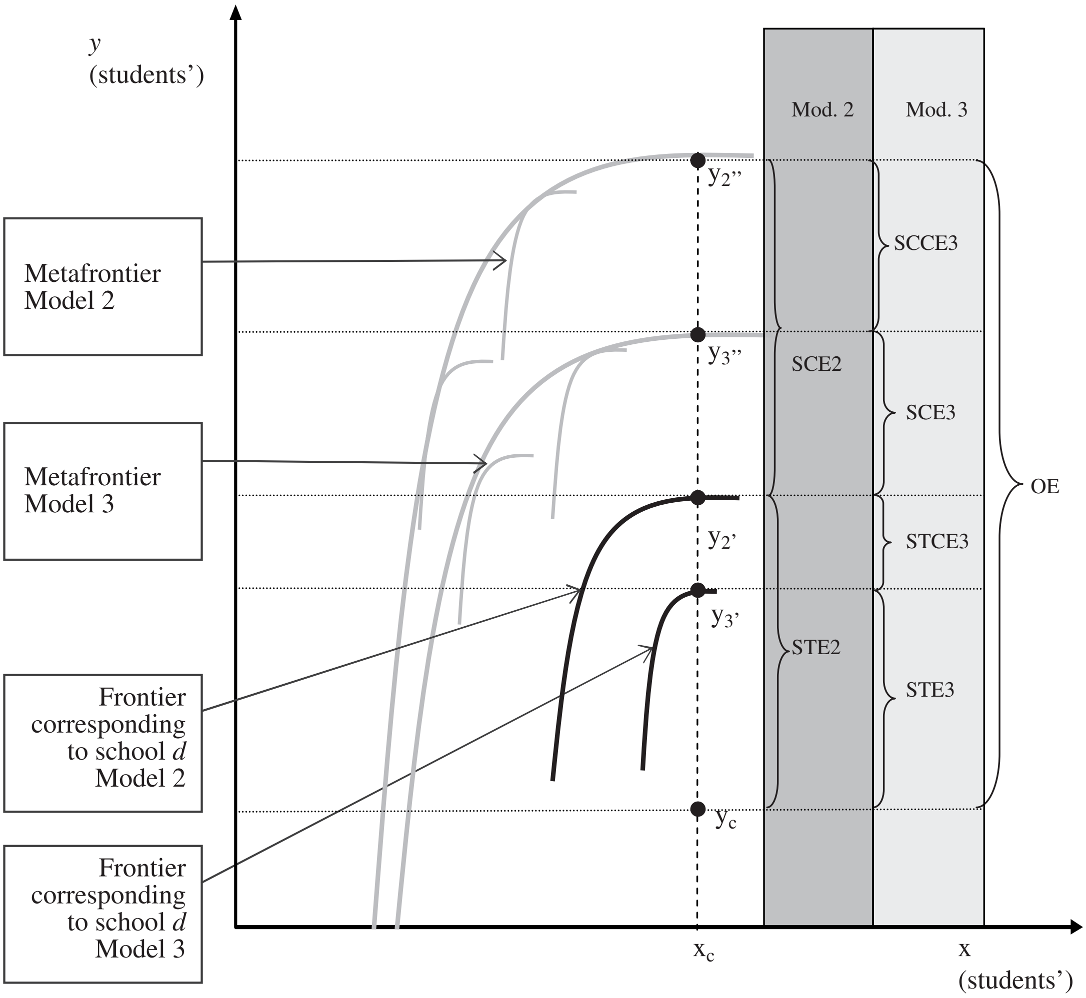
In contrast, our proposal here is based on the definition of a *contextualized value-added* robust multi-level nonparametric frontier assessment that separates the net effects of student and school, controlling for socioeconomic status, both at the student and school level, eliminating (or at least drastically reducing) the potential problems caused by the existence of outliers and dimensionality problems, as will be explained in the following section.

3. Methodology

3.1. The decomposition of overall efficiency

Following the rationale described in the above paragraphs, our model is inspired by Silva Portela and Thanassoulis's (2001) initial contribution, in which two frontiers are considered, namely, the local and the global frontiers. Whereas the former is specific to each school, and oriented to an estimation of student- within-school efficiency, the latter is used to estimate student within-all-schools efficiency. The so-called student's effect (henceforth *STE*), or student's efficiency, will determine the distance to the local frontier. In contrast, the school's effect (henceforth *SCE*), or school's efficiency, refers to the distance separating the local and the global frontiers. Model 2 in figure 1 documents the ideas underlying both effects.

FIGURE 1: **Descomposition of the *pure* value added (model 2) and *contextual* value added (model 3)**



Source: European Central Bank.

In figure 1, the student (c) achieves an output level represented by y_c, corresponding to an input level x_c —the *score* achieved by the student in a previous academic year. When the student's (c) academic performance is compared with the local frontier (which corresponds to the school in which student *c* is enrolled, i.e. school *d)*, one may notice that student *c* is inefficient. This occurs because on the frontier there are more efficient students enrolled in the same school who achieve better results (y'_2) using the same inputs —or previous knowledge (x_c). Therefore, the student's effect, or what Silva Portela and Thanassoulis (2001) refer to as the *student-within-school's efficiency*, is determined as the ratio of the potential to the actual output, i.e. $STE_2 = y'_2 / y_c$. This student's effect is higher than unity when the student is inefficient (as in the case presented in figure 1), and equal to unity otherwise. The efficiency coefficient for the student under analysis will be $OE_2 = y''_2 / y_c$ when compared to the overall frontier—metafrontier, or the student-within-all-schools' efficiency in the terms used by Silva Portela and Thanassoulis (2001). Having these two reference frontiers, the school's effect (SCE_2, a sort of technology-gap ratio separating the school-specific frontier from the overall frontier) is determined by comparing the overall and local frontiers ($SCE_2 = y''_2 / y'_2 = OE_2 / STE_2$).

In summary, the proposal of Silva Portela and Thanassoulis (2001) decomposes the global efficiency into two effects, namely:

$$\text{Overall efficiency } \left(OE_2\right) = \text{Student's effect} \left(STE_2\right) \times \text{School's effect} \left(SCE_2\right) \tag{1}$$

According to the taxonomy proposed in table 1, this decomposition corresponds to model 2, or *pure* value added —when the score indicating past knowledge achieved acts as an input. As mentioned above, we partially follow this proposal as we are interested in the student as the unit of analysis. However, in contrast to Silva Portela and Thanassoulis (2001), our aim is to develop a multilevel decomposition of *contextual* value added (CVA). This implies considering not only students' academic results, but also contextual factors regarding student and schools. To that end, we follow a previous proposal (Thieme et al. 2013), classifiable as a type 1 model (considering only the contextual status but not the value added; see table 1), which introduces successive decompositions after the consideration of specific variables. The final effect is the modification of the school effect, after introducing contextual variables on the average socio-economic level of the parents of students attending the same school.

In figure 1 we illustrate the differences between the proposal of Silva Portela and Thanassoulis (2001) which is, precisely, an example of model 2 (consisting of the assessment of the pure value added), and our proposal, which gives rise to model 3 (a CVA assessment). Regarding the student's effect, we see that $y'_2 > y'_3$ because in model 3 we consider as inputs not only the previous scores but also the socio-economic and cultural level of the student's family. This would imply that, in order to estimate y'_3, we consider the student's family context, whereas y'_2 implicitly assumes that this context is optimal and does not interfere with students' scores. In other words, benchmark y'_3 comes from a student that, having the same previous scores and comparable socio-economic situation, achieves a better academic outcome. In contrast, benchmark y'_2 is that corresponding to another student with a better socio-economic situation.

The contextual variables also have an impact on the school effect. Indeed, it is also clear that $y''_2 > y''_3$, since the contextual socio-economic environment could also affect the student's achievement in y''_3, but model 2 assumes that this factor has no effect and, therefore, y''_2 is perfectly attainable.

Summing up, when comparing models 2 and 3 we see that part of what is considered as student's inefficiency (or student's effect, STE_2), according to model 3 is attributable to both the effect of the contextual variables ($STCE_3$) and the net student's effect (STE_3). Analogously, the school's effect from model 2 (SCE_2) can be decomposed in order to account for the impact of the context due to socio-economic factors ($SCCE_3$) and the net effect of the school (SCE_3). This implies that a potential technology gap (represented by both $STCE_3$ and $SCCE_3$) appears when the context has a significant impact on the scores that students can achieve. In model 3 this gap may, or may not, be significant whereas in model 2, by definition, the impact is nonexistent. According to these arguments, the decomposition corresponding to model 3 can be expressed as:

$$\begin{aligned} \text{Overall efficiency} &= \text{Contextual effect on efficiency} \times \text{Net overall efficiency} \\ &= \text{Contextual effect on efficiency} \times \text{Student's effect} \times \text{School's effect} \end{aligned} \tag{2}$$

or:

$$OE = (y''_2 / y_c) = (y''_2 / y''_3) \times (y'_2 / y'_3) \times (y'_3 / y_c) \times (y''_3 / y'_2) \tag{3}$$

3.2. Using partial frontiers

The first decision to be made when estimating inefficiency levels is the specification of the technology, which has relevant implications —i.e. different technologies could lead to different results. Many previous applications have considered DEA, implying that a non-convex technology is assumed —i.e. each inefficient student will be compared to her more efficient peers, or combinations of them. In contrast, FDH requires comparison with an *existing* student, and linear combinations are not allowed as a benchmark —i.e. the convexity assumption is dropped.

Both DEA and FDH have some shortcomings, among which we may highlight the so-called curse of dimensionality, their lack of statistical properties (as they are deterministic in nature), and the potential impact of outliers. Some studies have established the statistical properties of the FDH estimator (Kneip et al. 1998; Simar and Wilson 2000), indicating that the dimensionality problems of the FDH models originate from their slow convergence rates. However, their statistical properties are very appealing, since they are consistent estimators for any monotone boundary —i.e. by imposing only strong disposability. In addition, as shown by Park et al. (2000), FDH has additional advantages over convex models, since the latter causes a specification error when the true technology is non-convex[3].

In our study we will assume non-convex technologies which, in our particular setting will imply that an *existing* student will be compared with another *existing* albeit more efficient student. However, FDH approaches have also some limitations and, therefore, we will consider a partial frontier approach such as order-*m* (Cazals et al. 2002), which is much more robust to both outliers and the curse of dimensionality[4]. Therefore, we will proceed in two stages, defining the FDH evaluation process in the first stage, and phasing in the order-*m* method in the second one. The details have been deferred to Appendix A.

4. Data, Inputs and Outputs

THE database used in this study was built from SIMCE data which since the mid-1990s, has been assessing, for the total pupil population, the learning processes of 4th and 8th grade primary education students, as well as 2nd grade secondary education students, via standardized tests. These tests evaluate the level of achievement of the fundamental objectives in these grades. From year 2006 onwards, 4th grade students (primary education) are evaluated every

[3] Some authors such as Thanassoulis and Silva Portela (2002) have proposed other methods to overcome the issue of outliers, by identifying and eliminating the extreme (super-efficient) cases; however, this is a controversial approach, since these units can convey relevant information. In the particular context of education, eliminating super-efficient observations could lead to an increase of overall efficiency —magnifying mediocrity, and reducing potential efficiency gains that could be achieved.

[4] In a recent paper, Krüger (2012) ranked the order-m estimation method as dominated, in general conditions, by the stochastic frontier, DEA and FDH methods. So, it appears that its use should be restricted to those cases characterized by the significant presence of outliers. This is precisely the case of education, where some students with very limited resources may achieve a brilliant academic curriculum.

year, whereas those of 8th grade and 2nd grade of secondary education are evaluated every other year. In addition to the tests associated to the curriculum, the SIMCE also includes information on teachers, students and parents using questionnaires on the context. This information is then used to contextualize and to analyze students' results in the SIMCE tests.

For this study we used information on results for 4th grade students, for year 2005, as well as 8th grade students, for year 2009. Similarly, information on families' socioeconomic and cultural level was also obtained via the questionnaire applied to the parents for year 2005 evaluation.

Accordingly, out of 142,109 students who took both exams, and were enrolled in the same school, only 57,000 also met the requirement of being enrolled in a school with more than 30 students taking the exam, and complete information existed for more than 60% of students who took the SIMCE exam in that school in 2009. A smaller sub-sample of 47,076 students in 948 schools was drawn from this large sample, out of which 395 (41.67%) were enrolled in municipal (public) schools, 460 (48.52%) were enrolled in privately-owned subsidized schools, and 93 (9.81%) were enrolled in either private non voucher schools or fee-charging private voucher schools. This information is reported in table 2.

The different models to evaluate are based on the availability of information gathered at student level (5 variables) and school level (1 variable). As output variables we consider the scores obtained by the 8th grade students in 2009 in the SIMCE tests in Mathematics and Language (y_1 and y_2, respectively). Regarding the input variables at student level, we consider two variables of previous academic achievement such as the score obtained by the same students in the SIMCE Mathematics and Language tests when they were in 4th grade in 2005 (x_1 and x_2, respectively). We also consider an index of the socioeconomic and cultural level of the families, constructed using principal components analysis and including the variables corresponding to both parents' educational level and the average family monthly income. This information was obtained using a questionnaire applied to families together with the SIMCE test during year 2005 (x_3). Analogously, as a proxy for the socioeconomic level of the school we used the average of the families' socioeconomic and cultural levels (x_4).

TABLE 2: **Sample description**

Type of school	Pupils		Schools	
	Number	Percentage	Number	Percentage
Public	18,021	38.28	395	41.67
Privately-owned subsidized	23,987	50.95	460	48.52
Privately-owned fee paying	5,068	10.77	93	9.81
Total	**47,076**	**100.00**	**948**	**100.00**

TABLE 3: **Summary statistics for the different variables**

Level	Variable	Description	#	Mean	Std. dev.
	y_1	Mathematics score, 8th grade, 2009	47,076	274.1052	51.4150
	y_2	Language score, 8th grade, 2009	47,076	263.9506	50.1096
	x_1	Mathematics score, 4th grade, 2005	47,076	267.0472	50.5391
	x_2	Language score, 4th grade, 2005	47,076	273.7696	48.4829
Student level	x_3	Socioeconomic and cultural level, student's family	47,076	0.2923	1.0111
School level	x_4	Socioeconomic and cultural level, school average	47,076	0.2917	0.8372

Following the classification provided in Section 2, we consider four evaluation models (one of which, model 0, corresponds to SIMCE average). The descriptive statistics of the variables included in the different models are reported in table 3.

5. Results

5.1. Overall performance and its decomposition: models and school types

Table 4 reports the results for the four types of evaluations: type 0 (status model without contextual variables), type 1 (status model with contextual variables), type 2 (pure value added model), type 3 (CVA model). The results of type 0 model correspond to the raw results of the average for each student, similar to those reported every year in Chile. The other three types of models correspond to the order-*m* results obtained for each of them.

TABLE 4: Results for the overall effect (OE) and its components, geometric means

	Type of effect	Model 0 (Status)	Model 1 (CS)	Model 2 (VA)	Model 3 (CVA)
Total	Overall effect (*OE*)	265.5373	1.1737	1.0545	1.0155
	Student effect (*STE*)	–	1.1515	1.0283	1.0179
	School effect (*SCE*)	–	1.0192	1.0255	0.9976
	Contextual effect (*CXTE*)	–	–	–	1.0385
Public schools	Overall effect (*OE*)	246.2879	1.2335	1.1019	1.0477
	Student effect (*STE*)	–	1.1678	1.0289	1.0177
	School effect (*SCE*)	–	1.0563	1.0710	1.0294
	Contextual effect (*CXTE*)	–	–	–	1.0518
Privately-owned subsidized	Overall effect (*OE*)	273.5235	1.1500	1.0347	1.0018
	Student effect (*STE*)	–	1.1479	1.0274	1.0176
	School effect (*SCE*)	–	1.0019	1.0071	0.9845
	Contextual effect (*CXTE*)	–	–	–	1.0329
Privately-owned fee-paying	Overall effect (*OE*)	315.7072	1.0508	0.9610	0.9513
	Student effect (*STE*)	–	1.1021	1.0304	1.0202
	School effect (*SCE*)	–	0.9534	0.9326	0.9325
	Contextual effect (*CXTE*)	–	–	–	1.0101

Results for the type 0 model are reported in the first column. They indicate remarkable differences in performance among school types. They also indicate that paid private schools outperform their subsidized counterparts, and these, in turn, have better performances than municipal schools. However, it should also be taken into account that this type of evaluation erroneously attributes all existing differences (overall effect) to the school (school effect). This occurs because, according to this type of evaluation, all students are assumed to have similar cognitive abilities as well as optimal input and environmental factors (socioeconomic and cultural level of the student's family).

However, a deeper scrutiny of these results shows an intraclass correlation of 34%. This result provides an indication of the dependence of the scores of student achievement on the school. About 34% of the variance end of student achievement can be attributed to the school. This implies that the greatest amount of variance occurs at the student level, which is not easily visualized when this type of results are examined.

The order-*m* results for the contextual status model (type 1), which decomposes the total effect between pupil effect and school effect are reported in the second column of table 4. This model considers observable (and therefore not optimal) values of uncontrollable factors at both student and school levels. For the total sample, overall inefficiency (overall effect) is 1.1737 (geometric mean), mainly dominated by the student effect, whose value is 1.1515, and to a much lesser extent by the school effect, which takes a value of 1.0192. Therefore, on average, the contribution of the student effect is much greater than that attributable to the school. The breakdown of these results by type of school indicates that for all three categories the student effect is much more relevant than the school effect on the overall inefficiency. In addition, similarly to what was found for the type 0 model, the overall differences across school types are also remarkable (the geometric means are 1.2335, 1.1500 and 1.0508 for public, privately owned subsidized and privately owned fee paying, respectively).

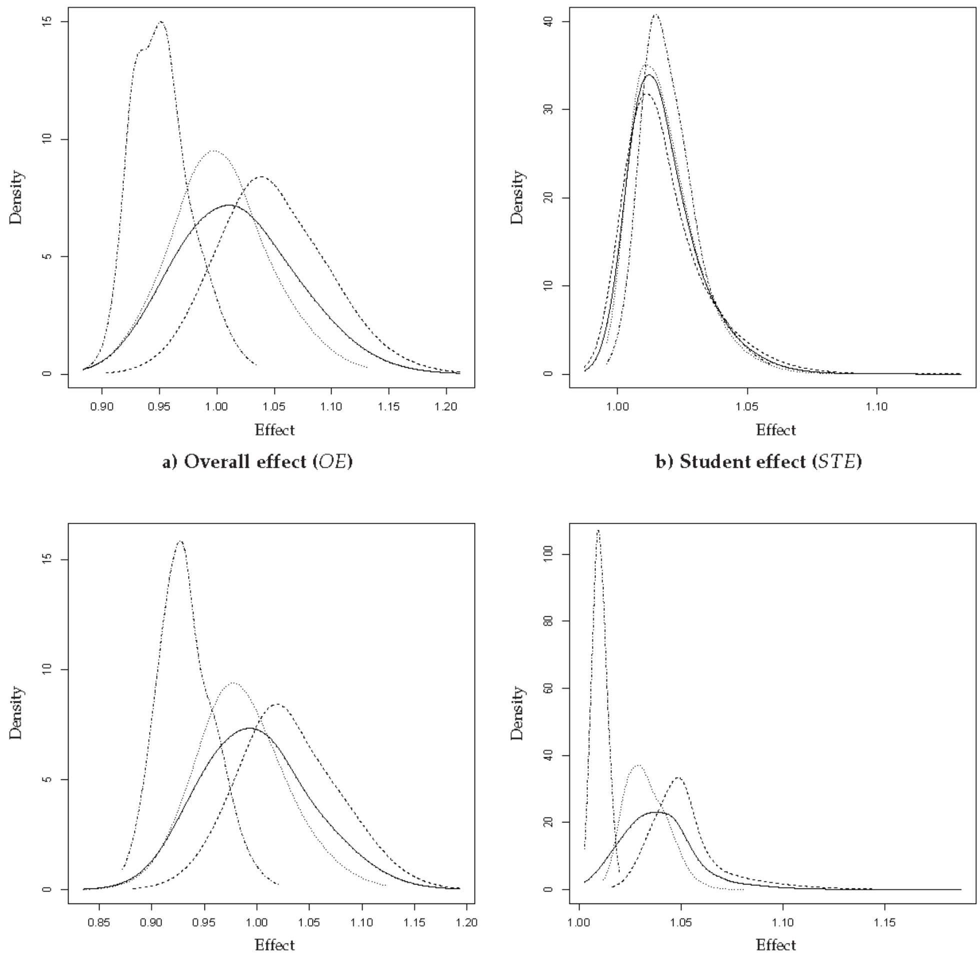
The results for the type 2 evaluation model (pure added value), reported in the third column of table 4, consider observed values of students' academic achievement at both the beginning and the end of the second cycle of basic education. However, it assumes that all students have the optimum deployment of environmental factors. For the whole sample, the geometric mean corresponding to the overall inefficiency (overall effect) is 1.0545, and the student effect (1.0283) is only slightly higher than the school effect (1.0255). However, in this case the decomposition of results by type of school is not analogous to that found either for the type 0 or for the type 1 model. Indeed, although the student effect is relatively stable across school types (between 1.0274 in the case of public schools and 1.0304 in the case of privately-owned fee paying), the school effect varies remarkably. In this case, on average (geometric mean), privately owned fee paying schools are the most efficient (0.9326), whereas public schools are the most inefficient (1.0710).

With respect to the more basic models, the type 3 model represents a value-added evaluation which goes further by incorporating contextual variables at both the student and school levels. The results for this type of evaluation (CVA model) are reported in the fourth model in table 4. For the whole sample, inefficiency decreases substantially with respect to the models evaluated in the preceding paragraphs, since the geometric mean is 1.0155. Indeed, much of this inefficiency is brought about by overall context variables (1.0385), followed by the student effect (1.0179), while the average effect at the school level is close to efficiency (0.9976).

Decomposing the results by type of school we obtain the same ranking as those obtained for the more basic models. Again, regarding the overall effect, public schools are the most inefficient, whereas privately-owned fee paying schools are the most efficient. Regarding the components of the school effect already taken into account (student and school effect), the sorting is also similar —on average. However, there is a now a third effect, the contextual effect (*CXTE*). This effect is clearly the worst for public schools (1.0518), whereas its impact is less pronounced for privately-owned subsidized schools (1.0329) and privately-owned fee paying schools (1.0101). Regarding the other two effects that had already been included in the previous models, in the case of the school effect only for public schools do we find inefficiency, on average. Both privately-owned subsidized and fee paying schools are, on average, efficient or, more specifically, super-efficient, since their corresponding geometric means are below one. In contrast, the student effect indicates there is inefficiency for three types of schools, and its magnitude is similar —on average.

A more detailed view of the results is gained by inspecting figure 2. Each sub-figure represents densities estimated using kernel smoothing for the different components of the type 3 model, and in each particular sub-figure the densities for all school types are also depicted. Therefore, figure 2 is the entire distribution counterpart to the fourth column in table 4, since the results reported are not restricted to some summary statistics such as the (geometric) mean or the variance (not reported). This might be important in some particular contexts, since neither the mean nor any dispersion indicator can detect if some scenarios such as polarization or multi-modality are emerging. Although using additional summary statistics may help in this regard, we consider it is much more informative to analyze the entire shape of the distributions, which provide much more encompassing views.

FIGURE 2: Kernel density plots by type of school, model 3



a) Overall effect (OE)

b) Student effect (STE)

c) School effect (SCE)

d) Contextual effect ($CXTE$)

All schools ——— Public schools ------- Privately-owned subsidized schools ········ Private-owned fee paying schools -·-·-·-·-

Note: All figures contain densities estimated using kernel density estimation for the different components of the bipartite decomposition in expression (3). The vertical lines in each plot represent the average for each component of the decomposition. Densities were estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel was chosen.

The results shown in figure 2 generally corroborate those reported in table 4. As indicated in figure 2.a, the most favorable overall effect (OE) is found for the privately-owned fee paying schools (dash-dotted line), whose density is tightly concentrated in the vicinity of 0.95. In contrast, the density for both public schools (dashed line) and privately-owned subsidized schools (dotted line) shifts right-wards, especially in the case of the former. However, in both cases the probability mass is much more spread than in the case of the privately-owned fee paying schools, generating a remarkable amount of overlapping between the overall effect for both types of schools. The overlapping found between privately-owned fee paying and public schools is much more modest, however.

These results are very similar for the school effect (*SCE*) reported in figure 2, which might be suggesting that the overall effect is mainly driven by the school effect. The contextual effect (*CXTE*), shown in figure 2.d, points in this direction as well. In this case, the differences are more pronounced, with no overlapping at all between the contextual effect for public schools (figure 2.d, dashed line) and that found for privately-owned fee paying schools (figure 2.d, dash-dotted line). In contrast, the student effect (*STE*), shown in figure 2.b, indicates that the differences for all school types are negligible.

We can test formally whether the visual differences observed among school types in figure 2 are actually statistically significant or not. The Li (1996) test provides the means to do so. This test, also based on kernel smoothing, compared to other nonparametric tests such as the Wilcoxon or Kruskal-Wallis tests has the great advantage of testing whether the differences between the entire distributions, say $f(\times)$ and $g(\times)$, are significant or not. Despite these great advantages, the applications are still very scarce (see, for instance Kumar and Russell 2002).

TABLE 5: **Distribution hypothesis tests (Li, 1996), Model 3 (CVA)**

		Overall effect *(OE)*	Student effect *(STE)*	School effect *(SCE)*	Contextual effect *(CXTE)*
Public vs. privately-owned subsidized	*t*-statistic	33.0617	0.6868	33.6032	58.7384
	p-value	0.0000	0.2461	0.0000	0.0000
Public vs. privately-owned fee paying	*t*-statistic	46.7628	3.1236	47.6590	54.6167
	p-value	0.0000	0.0009	0.0000	0.0000
Privately-owned subsidized vs. fee paying	*t*-statistic	25.5732	0.9497	24.9796	54.4474
	p-value	0.0000	0.1711	0.0000	0.0000

Results for the Li (1996) test indicate that the differences observed in figure 2 are mostly significant at the 1% level. Actually, even for the case with the greatest amount of overlapping (*STE*, figure 2.b), the differences between public and privately-owned fee paying schools were significant. Among all the available comparisons, only those corresponding to the student effect in figure 2.b in which the privately-owned subsidized schools were compared were not significant.

5.2. Comparisons across the different evaluation models

The first methodology we use to compare the different models considered is to estimate school correlation coefficients among the four types of models. Table 6 presents the results of the correlations for the overall effect (*OE*). These indicators give a first approximation to the question of how well the estimates of the different types of evaluation models provide equivalent results.

TABLE 6: **Correlations for the overall effect (OE) across models, all schools**

	Model 0 (Status)	Model 1 (CS)	Model 2 (VA)	Model 3 (CVA)
All schools				
Model 0 (Status)	–	0.9679	0.9473	0.8725
Model 1 (CS)		–	0.9456	0.9296
Model 2 (VA)			–	0.9585
Model 3 (CVA)				–
Public schools				
Model 0 (Status)	–	0.9545	0.9167	0.7980
Model 1 (CS)		–	0.9117	0.8806
Model 2 (VA)			–	0.9208
Model 3 (CVA)				–
Privately-owned subsidized schools				
Model 0 (Status)	–	0.9552	0.8967	0.8445
Model 1 (CS)		–	0.9086	0.9186
Model 2 (VA)			–	0.9759
Model 3 (CVA)				–
Privately-owned fee paying schools				
Model 0 (Status)	–	0.9819	0.8479	0.8470
Model 1 (CS)		–	0.8343	0.8568
Model 2 (VA)			–	0.9919
Model 3 (CVA)				–

The correlations found for the overall effect, as shown in table 6, are in general high (mostly above 0.8). The correlation between the results of CVA (model 3) and the other three models are high, with values ranging between 0.7980 (compared to Model 0 in the case of public schools) and 0.9919 (compared with model 2, *VA*, in the case of privately-owned fee paying schools). These results bear several resemblances to those reported in studies comparing various value-added models using hierarchical linear traditional methodologies (Timmermans et al. 2011; Gorard 2006, 2008).

The correlations among the different models generally offer useful information. However, they have the same disadvantage as that attributable to the information contained in table 4, i.e. some important phenomena are boiled down to a summary statistic —in table 4 it is the mean, in table 6 it is the correlation coefficient. Yet the way in which two given distributions may be too complex to be summarized in a single statistic only —i.e. the law of movement between one distribution (say, model 2) and another one (say, model 3) should be modeled more carefully.

Accordingly, table 7 provides results on transition probabilities across models 0 (status) and 3 (con-textual value added), and across models 2 (value added) and 3. As indicated by the results in table 7.a, the comparison of the models 0 and 3 indicates that, out of the 100% of schools classified in the highest quintile (*Q*5) according to the average SIMCE (type 0), 69% remain in the same quintile according to an analysis of value added, 22% moved to quintile 4, 8% to quintile 3 and 1% to quintile 2. This result is contained in the last row of the 5 × 5 matrix in table 7.a. In contrast, as indicated in the first row in the same 5 × 5 matrix, out of the 100% of schools classified in the lowest quintile (*Q*1, 20% in line with the average worst SIMCE performance), 69% would be in the same classification according to a contextual value-added assessment, 26% should be classified in quintile 2, 4% in quintile 3 and 1% in quintile 4. In the case of the central quintiles, the intra-distribution mobility is much higher, as shown by the rest of the entries on the main diagonal (0.39, 0.40 and 0.45 for quintiles 2, 3 and 4, respectively).

The transitions across models 2 and 3 are shown in table 7.b. In this case, the entries on the main diagonal average to 0.71, compared with the 0.52 of table 7.a, indicating lower intra-distribution mobility— i.e. the different schools are more stable in their rankings when contextual variables are introduced. However, on average, almost 30% of the schools move in this ranking, suggesting the importance of including this type of effect when modeling school effectiveness. In the case of the central quintiles of the distributions, the changes in the relative positions are still higher (only 64%, on average, remain in their relative positions), and even higher when modeling the transitions between model 0 and model 3 (barely 41% of the central quintiles remain in their relative positions).

TABLE 7: **Transitions across models**

a) Model 0 to 3

		Upper quintile, model 3					(Number)
		Q1	**Q2**	**Q3**	**Q4**	**Q5**	
Upper quintile, model 0	Q1	0.69	0.26	0.04	0.01	0.00	(190)
	Q2	0.28	0.39	0.29	0.05	0.00	(189)
	Q3	0.02	0.25	0.40	0.28	0.05	(190)
	Q4	0.01	0.10	0.19	0.45	0.25	(189)
	Q5	0.00	0.01	0.08	0.22	0.69	(190)

b) Model 2 to 3

		Upper quintile, model 3					(Number)
		Q1	**Q2**	**Q3**	**Q4**	**Q5**	
Upper quintile, model 2	Q1	0.83	0.17	0.01	0.00	0.00	(190)
	Q2	0.17	0.65	0.18	0.00	0.00	(189)
	Q3	0.00	0.16	0.63	0.20	0.01	(190)
	Q4	0.00	0.02	0.15	0.64	0.19	(189)
	Q5	0.00	0.00	0.03	0.16	0.81	(190)

6. Conclusions

THIS paper has attempted to make a twofold contribution —both methodological and empirical— to the evaluation of school performance. Regarding the methodology considered, our study differs in two aspects from the existing proposals in the literature: the techniques used for performance evaluation and the type of model considered. With respect to our proposed methodological approach for evaluating school effectiveness, unlike most studies in this area which use multilevel regression approaches to measure the net contribution of the student and school in the student's achievement, we use activity analysis techniques. In this regard, similarly to the recent contributions by De Witte et al. (2010) and Thieme et al. (2013), we use order-*m* techniques to alleviate the problems of dimensionality and the influence of outliers, obtaining more statistically robust results.

With respect to the type of model considered, previous proposals such as Silva Portela and Thanassoulis (2001) or De Witte et al. (2010) considered only variables of academic achievement at the *student* level. Instead, we consider contextual variables at both the student and the school levels. This approach is more in line with the studies in the economics of education literature, and enables us to propose a model of CVA within the methodological context of activity analysis techniques. To our knowledge, no previous studies have been published using methodologies to study school performance.

Indeed, the relevant literature of value added and multilevel analysis in education, as well as the results found in this paper, show how important and necessary it is to include contextual variables that are beyond the school's control (the most notable example is students' socioeconomic status) at different levels under analysis, mainly for two reasons. First, if we do not consider contextual variables when estimating sub-performing schools that students with unfavorable economic conditions attend, it could result in managers blaming these centers for variables that are beyond their control or, even worse, it might create perverse incentives to select students according to their socio-economic characteristics.

Second, our results also indicate, corroborating the existing literature, that performance differences among the different types of institutions (privately-owned fee paying, privately-owned subsidized, and public schools) decrease significantly when controlling for these variables. From an empirical point of view, the results show that most of the large differences observed in the raw results of academic achievement among types of schools are largely explained away when incorporating variables that account for students' entry conditions and contextual variables. Thus, much of the variance explained by the school, which is particularly high in Chile due to the high level of school segregation in the country, disappears once these variables enter the model. Indeed, the study results show that, for the total sample average, a *pure* value added model (type 2 model, which does not consider contextual variables), allocates approximately 53% of overall inefficiency to the student effect and the remaining 47% to the school effect. However, a contextual VA model (type 3 model) attributes only 33% of the inefficiency to the student effect and the remaining inefficiency to the student contextual effect (individual/family and school). So, clearly a VA model over evaluates pure inefficiency of schools whose students do not have optimal socioeconomic status.

With respect to previous literature, the results are in agreement with it. In particular, it can be observed that: (i) there is a high correlation between the CVA model (model 3) with the academic achievement *raw* results, although this represents an important change in the ranking of schools according to their performance; (ii) the educational system mimics and even deepens the existing social inequalities: the highest the socioeconomic level, the better schools can be accessed to, and the lower the negative effect of the environment; (iii) the high discrepancies in learning skills among students with varying socioeconomic levels is particularly strong in the first school years; however, when isolating the *pure* student effect, there are no differences among students in different types of schools.

This situation becomes more visible when analyzing the results according to the type of schools. Due to the characteristics of the Chilean education system, there is a close correlation between the type of school and the socioeconomic status of their students. Indeed, for privately-owned fee-paying schools, catering to the more affluent sectors of society, and therefore with contextual variables very close to optimal for the vast majority of their students, the differences between the results for type 2 and type 3 models do not vary significantly. In fact, the school effect remains almost unchanged. In contrast, for public schools, catering to the lower income population, the results of school effect change significantly. In this case, the contextual effect incidence almost halves that corresponding to the overall inefficiency.

Appendix A. Using order-*m* in a contextual value-added model

ACCORDINGLY, we assume there is information available on the input and output vectors ($xc = (x_{c,1}, x_{c,2}, \ldots, x_{c,i}, \ldots, x_{c,I})$ and $yc = (y_{c,1}, y_{c,2}, \ldots, y_{c,j}, \ldots, y_{c,J})$, respectively) for each student in the sample (1, 2, …, C). We will then characterize the elements of the integer activity vector as $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_C)$ and the efficiency coefficient as α_c^{FDH} .

Then, the output-oriented FDH efficiency scores will be yielded by solving the following linear programming problem:

$$
\begin{aligned}
&\left\{\alpha_c^{FDH\max_{\lambda_1,\lambda_2,\ldots,\lambda_C}}\right\}^{\alpha_c^{FDH}} && \\
&s.t. && \\
&\sum_{s=1}^{C} \lambda_S x_{s,i} - x_{c,i} \geq 0, && i = 1,\ldots,I \\
&-\sum_{s=1}^{C} \lambda_S y_{s,j} + \alpha_c^{FDH} y_{c,j} \geq 0, && j = 1,\ldots,J \\
&\sum_{s=1}^{C} \lambda_S = 1 && \\
&\lambda_S \in \{0,1\}, && s = 1,\ldots,S
\end{aligned}
\tag{4}
$$

Linear programming problem (4) identifies, for each student c to be FDH-efficient, another student in the sample with better performance —i.e. the student with coefficient $\lambda_{s*} = 1$. Then it estimates the output increase $1-\alpha_c^{FDH}$ which is needed to reach the non-convex frontier, $\alpha_c^{FDH} > 1$. Therefore, by solving linear programming problem (4), we will have an activity vector $\lambda_c = 1$ and an efficiency coefficient $\alpha_c^{FDH} = 1$ for FDH-efficient students.

As indicated earlier, relatively recent contributions in the literature provide methods to overcome the curse of dimensionality and the effect of outliers inherent to FDH. Among them, the order-*m* estimator (Cazals et al. 2002; Simar 2003) has become one of the most popular methods to get round these issues while at the same time maintaining the advantages of a non-convex and nonparametric methodology.

According to this method, we will first consider a positive fixed integer, m. For a given level of input ($x_{c,i}$) and output ($y_{c,j}$), the order-*m* estimation defines the *expected* value of maximum of m random variables ($y_{1,j},\ldots, y_{m,j}$), drawn from the conditional distribution of the output matrix $\mathbf{Y}$ for which $y_{m,j} > y_{c,j}$. Formally, the proposed algorithm to compute the order-*m* estimator has four steps:

1) For a given level of $y_{c,j}$, draw a random sample of size m with replacement among those $y_{m,j}$, such that $y_{m,j} \geq y_{c,j}$.
2) Compute program (4) and estimate $\tilde{\alpha}_c$.
3) Repeat steps 1 and 2 B times and obtain B efficiency coefficients $\tilde{\alpha}_c^b$ ($b = 1, 2, \ldots, B$). The quality of the approximation can be tuned by increasing B (in most applications $B = 200$ seems to be a reasonable choice, but we decided to set $B = 2000$).
4) Compute the empirical mean of B samples as:

$$\alpha_c^m = \frac{\sum_{b=1}^{B} \tilde{\alpha}_c^b}{B} \tag{5}$$

The number of observations considered in the estimation approaches the observed units that meet the condition $y_{m,j} > y_{c,j}$ as m increases, whereas the expected order-*m* estimator in each of the b iterations $\tilde{\alpha}_c^b$ tends to the FDH efficiency coefficient $\tilde{\alpha}_c^{FDH}$. Therefore, m is an arbitrary positive integer value, but it is always convenient to observe the fluctuations of the $\tilde{\alpha}_c^b$ coefficients that will ultimately depend on the level of m. α_c^m will normally take values higher than the unity for acceptable values of m (this indicates that these units are inefficient, as outputs can be increased without modifying the inputs allocated). When $\alpha_c^m < 1$, the unit c may be labeled as super-efficient, provided the order-m frontier shows lower c output levels than the unit under analysis.

In addition, as mentioned earlier, the order-m method is also an excellent tool for overcoming the dimensionality problems, as well as the presence of extreme observations or outliers. Yet in our particular setting the usefulness of the proposed evaluation can be limited if the inefficiency achieved is partly attributable to contextual factors which we do not wish to introduce in the assessment.

In order to refine the evaluation process, taking this into account, and as previously discussed in Model 3, we define a multilevel frontier assessment process that could be considered to estimate the impact of the socio-economic factors on students' efficiencies. In order to carry out this multilevel estimation, we adapt the metafrontier approaches proposed by Battese and Rao (2002), Battese et al. (2004) and O'Donnell et al. (2008). In the case of Model 3, this process has the following steps:

a) Classify students (1, 2, …, C) depending on the school they are enrolled in (1, 2, …, D).
b) Complete steps 1 to 4 to estimate the efficiency coefficients corresponding to each student in the specific school she/he is enrolled in (α_c^m) (meaning, considering the school frontier point represented by y'_3 in figure 1 in order to estimate STE). In order to facilitate the cross-comparison of the results, irrespective of the number of students classified in each school, the same value of m will be assigned in all the estimations. This neutralizes the problems of dimensionality and the potential impact of the outliers.
c) After completing the conditional frontiers, add new input variables (the socio-economic and cultural level corresponding to the student's family and the average of the same variable for the school) and apply again steps 1 to 4 of the order-m estimation to the complete sample to estimate the efficiency coefficients with respect to the metafrontier ($\alpha_{c,1}^m$). These new coefficients provide an assessment of the student's efficiency with respect to the overall metafrontier, taking into account only schools operating with no better environmental factors than the school where the student is enrolled (precisely what is represented by point y''_3 in figure 1).

For each student found to be FDH-inefficient, program (4) identifies another student in the sample with superior performance in order to estimate the increase in the output required to reach the non-convex frontier ($\alpha'_3 > 1$), where ($1-\alpha'_3 > 1$) the required proportional increase in the output level. For students declared FDH-efficient, program (4) offers an activity vector (λ_c) and an efficiency coefficient equal to the unity ($\lambda_c = 1;\ \alpha'_3 = 1$).

7. References

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

Badunenko, O., D.J. Henderson, and S.C. Kumbhakar (2012): "When, where and how to perform efficiency estimation". *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 175 (4), 863-892.

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

Battauz, M., R. Bellio and E. Gori (2011): "Covariate measurement error adjustment for multilevel models with application to educational data". *Journal of Educational and Behavioral Statistics* 36 (3), 283-306.

Battese, G.E. and D. Rao (2002): "Technology gap, efficiency, and a stochastic metafrontier function". *International Journal of Business* 1 (2), 87-93.

Battese, G.E., D. Rao and C.J. O'Donnell (2004): "A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies". *Journal of Productivity Analysis* 21 (1), 91-103.

Bessent, A., W. Bessent, J. Kennington, and B. Reagan (1982): "An application of mathematical programming to assess productivity in the Houston independent school district". *Management Science* 28 (12), 1355-1367.

Blanco, E. (2010): "Eficacia escolar y desigualdad: aportes para la política educativa". *Perfiles Latinoamericanos* 17(34):

Bock, R., F. Wolfe and T. Fisher (1996): *A Review and Analysis of the Tennessee Value-Added Assessment System*. Nashville: Tennessee Comptroller of the Treasury, Office of Educational Accountability.

Bryk, A.S. and S.W. Raudenbush (1992): *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Carlson, D. (2001): *Focusing state educational accountability systems: four methods for judging school quality and progress*. Technical report, Dover, NH: Center for Assessment (NCIEA).

Carnoy, M. and S. Loeb (2002): "Does external accountability affect student outcomes? A cross-state analysis". *Educational Evaluation and Policy Analysis* 24 (4), 305-331.

Cazals, C., J.P. Florens and L. Simar (2002): "Nonparametric frontier estimation: a robust approach". *Journal of Econometrics* 106, 1-25.

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

Coleman, J.S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F.D. Weinfeld and R.L. York (1966): *Equality of Educational Opportunity*. Washington, DC: US Department of Health, Education, and Welfare, Office of Education.

De Witte, K., E. Thanassoulis, G. Simpson, G. Battisti and A. Charlesworth-May (2010): "Assessing pupil and school performance by non-parametric and parametric techniques". *Journal of the Operational Research Society* 61 (8): 1224-1237.

Denvir, B. and M. Brown (1986): "Understanding of number concepts in low attaining 7-9 year olds: Part I. Development of descriptive framework and diagnostic instrument". *Educational Studies in Mathematics* 17 (1), 15-36.

Drury, D. and H.C. Doran (2003): "The value of value-added analysis". *National School Boards Association Policy Research Brief* 3 (1), 1-4.

Ercikan, K. (2006): "Examining guidelines for developing accurate proficiency level scores". *Canadian Journal of Education* 29 (3), 823-838.

Goldstein, H. (1995): *Multilevel Statistical Models*. London: Wiley.

——— (2003): *Statistical Analysis of Multilevel Data*. Edward Arnold, London.

Goldstein, H., J. Rasbash, M. Yang, G. Woodhouse, H. Pan, D. Nuttall and S. Thomas, (1993): "A multilevel analysis of school examination results". *Oxford Review of Education* 19 (4), 425-433.

Goldstein, H. and S. Thomas (1996): "Using examination results as indicators of school and college performance". *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159, (1), 149-163.

Gorard, S. (2006): "Value-added is of little value". *Journal of Educational Policy* 21 (2), 233-241.

——— (2008): "The value-added of primary schools: what is it really measuring?". *Educational Review* 60 (2), 179-185.

Gray, J., D. Jesson, H. Goldstein, K. Hedger and J. Rasbash (1995): "A multi-level analysis of school improvement: Changes in schools' performance over time". *School Effectiveness and School Improvement* 6 (2), 97-114.

Hanushek, E.A. and M.E. Raymond (2005): "Does school accountability lead to improved student performance?". *Journal of Policy Analysis and Management* 24 (2), 297-327.

Hox, J.J. (2002): *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kane, T.J. and D.O. Staiger (2002): "The promise and pitfalls of using imprecise school accountability measures." *The Journal of Economic Perspectives* 16 (4), 91-114.

Kneip, A., B.U. Park and L. Simar (1998): "A note on the convergence of nonparametric DEA estimators for production efficiency scores". *Econometric Theory* 14, 783-793.

Krüger, J.J. (2012): "A Monte Carlo study of old and new frontier methods for efficiency measurement". *European Journal of Operational Research* 222, 137-148.

Kumar, S. and R.R. Russell (2002): "Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence". *American Economic Review* 92 (3), 527-548.

Kupermintz, H. (2003): "Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System". *Educational Evaluation and Policy Analysis* 25, 287-298.

Li, Q. (1996): "Nonparametric testing of closeness between two unknown distribution functions". *Econometric Reviews* 15, 261-274.

Loader, C.R. (1996): "Local likelihood density estimation". *The Annals of Statistics* 24 (4), 1602-1618.

Mancebón, M.J. and C. Mar Molinero (2000): "Performance in primary schools." *Journal of the Operational Research Society* 51 (7), 843-854.

McCaffrey, D.F., J.R. Lockwood, D. Koretz, T.A. Louis and L. Hamilton (2004): "Models for value-added modeling of teacher effects." *Journal of Educational and Behavioral Statistics* 29 (1), 67-101.

McCaffrey, D.F., J.R. Lockwood, D.M. Koretz and L.S. Hamilton (2003): *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA: The RAND Corporation.

McEwan, P.J. (2003): "Peer effects on student achievement: Evidence from Chile". *Economics of Education Review* 22 (2), 131-141.

Meyer, R.H. (1997): "Value-added indicators of school performance: A primer". *Economics of Education Review*, 16(3):283-301.

Mizala, A. and P. Romaguera (2000): "School performance and choice: the Chilean experience." *Journal of Human Resources* 35 (2), 392-417.

Mizala, A., P. Romaguera and D. Farren (2002): "The technical efficiency of schools in Chile." *Applied Economics* 34 (12), 1533-1552.

Mortimore, P., P. Sammons and S. Thomas (1994): "School effectiveness and value added measures". *Assessment in Education* 1 (3), 315-332.

O'Donnell, C.J., D.S. Prasada Rao and G.E. Battese (2008): "Metafrontier frameworks for the study of firm-level efficiencies and technology ratios". *Empirical Economics* 34, 231-255.

Ouellette, P. and V. Vierstraete (2005): "An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method". *Applied Economics* 37 (14), 1643-1653.

Park, B.U., L. Simar and C. Weiner (2000): "The FDH estimator for productivity efficiency scores". *Econometric Theory* 16 (6), 855-877.

Portela, M.C., A.S. Camanho and A. Keshvari (2012): "Assessing the evolution of school performance and value-added: trends over four years". *Journal of Productivity Analysis*, forthcoming.

Raudenbush, S.W. (2004): *Schooling, Statistics, and Poverty: Can We Measure School Improvement?*. Princeton, NJ: Educational Testing Service.

Raudenbush, S.W. and A.S. Bryk (2002): *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Rouse, C.E., J. Hannaway, D. Goldhaber and D. Figlio (2007): "Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure". Working Paper No. 13681, National Bureau of Economic Research.

Ruggiero, J., W. Duncombe and J. Miner (1995): "On the measurement and causes of technical inefficiency in local public services: With an application to public education". *Journal of Public Administration Research and Theory* 5 (4), 403-428.

Sammons, P. (1995): "Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years". *British Educational Research Journal* 21 (4), 465-485.

Silva Portela, M.C.A. and E. Thanassoulis (2001): "Decomposing school and school-type efficiency". *European Journal of Operational Research* 132 (2), 357-373.

Simar, L. (2003): "Detecting outliers in frontier models: A simple approach." *Journal of Productivity Analysis* 20 (3), 391-424.

Simar, L. and P.W. Wilson (2000): "Statistical inference in nonparametric frontier models: The state of the art". *Journal of Productivity Analysis* 13 (1), 49-78.

Teddlie, C. and D. Reynolds (2000): *The International Handbook of School Effectiveness Research*. London: Routledge.

Tekwe, C.D., R.L. Carter, C.X. Ma, J. Algina, M.E. Lucas, J. Roth, M. Ariet, T. Fisher and M.B. Resnick (2004): "An empirical comparison of statistical models for value-added assessment of school performance." *Journal of Educational and Behavioral Statistics* 29 (1), 11-36.

THANASSOULIS, E. and M.C.A. SILVA PORTELA (2002): “School Outcomes: Sharing the Responsibility Between Pupil and School”. *Education Economics* 10 (2), 183-207.

THIEME, C., D. PRIOR and E. TORTOSA-AUSINA (2013): “A mutilevel decomposition of school performance using robust nonparametric frontier techniques”. *Economics of Education Review* 32, 104-121.

TIMMERMANS, A.C., S. DOOLAARD and I. DE WOLF (2011): “Conceptual and empirical differences among various value-added models for accountability.” *School Effectiveness and School Improvement* 22 (4), 393-413.

WILSON, D. (2004): “Which ranking? The impact of a value-added measure of secondary school performance”. *Public Money & Management* 24 (1), 37-45.

WILSON, D. and A. PIEBALGA (2008): “Performance measures, ranking and parental choice: an analysis of the English school league tables”. *International Public Management Journal* 11 (3), 344-366.

NOTA SOBRE LOS AUTORES - *ABOUT THE AUTHORS**

RENÉ GEMPP is a psychologist and doctoral student. He is currently is assistant professor in economics and business at Diego Portales University (Chile). His main research area is psychometrics, including scaling techniques, educational measurement, structural equation modeling and item response theory. He also has served as a psychometrician for SIMCE, the national standardized test administered annually in primary and secondary schools in Chile.
E-mail: rene.gempp@udp.cl

DIEGO PRIOR JIMÉNEZ, obtained his PhD from the Autonomous University of Barcelona where he is currently professor in finance and accounting and researcher at IESEG School of Management (Lille, France). He has been visiting lecturer at Diego Portales University (Santiago, Chile), Monterrey Institute of Technology (Guadalajara, Mexico), Kent School of Business (University of Kent, UK), and the Department of Economics and Related Studies (University of York, UK). Author of numerous books and monographs, he has published more than seventy articles in national and international journals such as *European Economic Review, Journal of Banking and Finance, European Journal of Operational Research, Journal of Productivity Analysis, OMEGA-The International Journal of Management Science*, etc. He has participated as associate researcher and main researcher on various national and international projects. Currently he is in charge of the project *Gestión y evaluación de las organizaciones* (Spanish Ministry of Science and Innovation, VI National Scientific Research, Development and Technological Innovation Plan 2008-2011).
E-mail: diego.prior@uab.cat

Any comments on the contents of this paper can be addressed to Diego Prior, Department of Business Economics, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain. Tel.: +(34) 935811539, fax: +(34) 935812555, e-mail: diego.prior@uab.

* This study was primarily supported by Grant #1121164 awarded to Claudio Thiem, René Gempp and Emili Tortosa-Ausina by FONDECYT (National Fund of Scientific and Technological Development) to research school value added in Chile. Diego Prior and Emili Tortosa-Ausina acknowledge the financial support of Ministerio de Ciencia e Innovación (ECO2010- 18967/ECON and ECO2011-27227). Emili Tortosa-Ausina also acknowledges the financial support of Generalitat Valenciana (PROMETEO/2009/066). All four authors are grateful for the comments by José Manuel Cordero as well as other participants at the XX Encuentro de Economía Pública (Seville, January 2013). The usual disclaimer applies.

CLAUDIO THIEME is a civil industrial engineer with a PhD in business administration from the Autonomous University of Barcelona. Currently he is associate professor in economics and business at Diego Portales University. He has carried out research studies in public management and policies for national ministries and local and regionalgovernments. His main research areas include organizational performance assessment and management processes, especially related to public and educational institutions.
E-mail: claudio.thieme@udp.cl

EMILI TORTOSA-AUSINA graduated in economics from the University of Valencia and holds a PhD in economics from Jaume I University, where he is currently a lecturer in applied economics. He has also taught at the University of Alicante (1994-1995) and has received scholarships from various institutions (Fundación Caja Madrid, among others). He has held posts as visiting researcher at the Autonomous University of Barcelona, the University of New South Wales (Australia), and the Department of Economics of Oregon State University (USA). His specialized fields are banking economics and the analysis of productivity and efficiency, application of different techniques applied to problems in the banking sector. He has published various books in collaboration with others and his articles have appeared in specialized journals. He has participated in many Spanish and international congresses and is the main researcher on the Spanish National Research Program *Evaluación de la eficiencia y la productividad de las organizaciones en un contexto de crisis económica: instituciones financieras y sector public.*
E-mail: tortosa@uji.es

ÚLTIMOS NÚMEROS PUBLICADOS – *RECENT PAPERS*

Documentos de Trabajo

5

Fundación BBVA

Plaza de San Nicolás, 4
48005 Bilbao
España
Tel.: +34 94 487 52 52
Fax: +34 94 424 46 21

Paseo de Recoletos, 10
28001 Madrid
España
Tel.: +34 91 374 54 00
Fax: +34 91 374 85 22

publicaciones@fbbva.es
www.fbbva.es